(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/45498 A2

(51) International Patent Classification[7]: A01K 67/027

(21) International Application Number: PCT/US01/46870

(22) International Filing Date: 5 December 2001 (05.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/254,302	8 December 2000 (08.12.2000)	US
60/255,228	11 December 2000 (11.12.2000)	US
60/256,191	13 December 2000 (13.12.2000)	US
60/256,197	13 December 2000 (13.12.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/254,302 (CON)
Filed on	8 December 2000 (08.12.2000)
US	60/255,228 (CON)
Filed on	11 December 2000 (11.12.2000)
US	60/256,191 (CON)
Filed on	13 December 2000 (13.12.2000)
US	60/256,197 (CON)
Filed on	13 December 2000 (13.12.2000)

(71) Applicant *(for all designated States except US)*: **DELTAGEN, INC.** [US/US]; 740 Bay Road, Redwood City, CA 94063 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALLEN, Keith, D.** [US/US]; 224 Custer Trail, Cary, NC 27513 (US). **BRENNAN, Thomas, J.** [US/US]; 325 Rockwood Drive, South San Francisco, CA 94080 (US).

(74) Agents: **BURKE, John, E.** et al.; Deltagen, Inc., 740 Bay Road, Redwood City, CA 94063 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSGENIC MICE CONTAINING CHANNEL PROTEIN TARGET GENE DISRUPTIONS

(57) Abstract: The present invention relates to transgenic animals, as well as compositions and methods relating to the characteristics of gene function. Specifically, the present invention provides transgenic mice comprising mutation in a target gene. Such transgenic mice are useful as models for disease and for identifying agents that modulate gene expression and gene function, and as potential treatments for various disease states and disease conditions.

WO 02/45498 A2

TRANSGENIC MICE CONTAINING CHANNEL PROTEIN TARGET GENE DISRUPTIONS

Field of the Invention

The present invention relates to transgenic animals, compositions and methods relating to the characterization of gene function.

Background of the Invention

Experimental animal models are important tools for understanding the role of genes. More particularly, the ability to develop animals with specific genes altered or inactivated has been invaluable to the study of gene function, and has lead to unexpected discoveries of a gene and/or mechanisms responsible for disease with similar manifestations in humans. These genetically engineered animals are also useful for identifying and testing therapeutic treatments for a variety of diseases and disorders.

The identification of the function of numerous genes has been useful in ascertaining their roles in disease. Because of the high level of homology between humans and mice, it is possible to define the function of individual human genes by making targeted germline mutations in selected genes in the mouse. The phenotype(s) of the resulting mutant mice can be used to help define the function of such genes in humans.

Identifying the roles of genes and their expression products may permit the definition of disease pathways and the identification of diagnostically and therapeutically useful targets. Of particular interest are cell membrane-spanning proteins and the genes encoding those proteins.

The cell membrane serves as a barrier to selectively keep molecules inside the cell or, conversely, keep molecules out of the cell. Whether or not molecules are allowed to cross this barrier depends on the needs of the cell. Raw materials needed for the cell to live are allowed to pass in, while waste materials that would eventually kill the cell are allowed to leave. This is how the cell membrane is responsible for controlling the internal environment of the cell. The cell membrane's structure is a lipid bilayer made up of phospholipids. The interior nonpolar region of the membrane forms a barrier to polar molecules. Since most of the food molecules, and water, are polar molecules, they pass into the cell through gateways provided by membrane proteins.

There are three types of membrane proteins that can be found imbedded in the cell membrane. They are channel proteins, receptor proteins, and marker proteins. Channel proteins allow specific materials to pass through the membrane. Specifically, a glucose channel protein, for example, will not allow water in, only glucose.

Ion channels are among the most important channel proteins in living organisms. These channels are the most fundamental elements of molecular hardware in the nervous system. They are the membrane-spanning proteins that directly mediate the transmembrane ionic fluxes giving rise to the generation, propagation, and integration of electrical signals in neurons, muscle, and other

electrically interesting cells. By forming aqueous pores right through the heart of the channel protein (and hence across the membrane the protein spans), channels act as "leakage" pathways for ions down their pre-established thermodynamic gradients. Channels discriminate fiercely among the different species of inorganic ions present in the aqueous solutions bathing the cell membrane. They also rapidly open and close their conduction pores in response to physiological signals, such as binding of neurotransmitters or changes in transmembrane electric field. Examples of important ion channels are those for regulating potassium, sodium and calcium ions.

Potassium ion (K^+) channels are ubiquitous membrane proteins responsible for the maintenance of the resting membrane potential and for the propagation of the action potential. Sequence analysis has identified two predominant types of K^+-channels: voltage-gated channels and inward-rectifier channels. Voltage-gated channels are identified by having six proposed transmembrane alpha-helices per subunit (S1-S6). Of these, S4, a highly charged segment, is believed to be the primary voltage-sensor. The inward-rectifier potassium ion channels are simpler in topology, having two membrane-spanning helices (M1 and M2) per subunit.

An expressed sequence tag (EST) (EST name: vt88g01.r1; GenBank Acc: AA718873; GenBank GI: 2731147) was isolated bearing sequence similarity and homology to Shab potassium ion channels. A need in the art exists to identify and characterize genes comprising this sequence (or a homog or ortholog thereof), which may play a role in dysfunctions or diseases.

Calcium ion channels are another particularly important subfamily of ion channels. Recently, the genes encoding the alpha 1 and beta subunits of voltage-sensitive calcium channel were mapped in the mouse by analysis of the progeny of two multilocus crosses (Chin *et al.*, 1995, *Genomics* 28(3):592-595). The alpha 1, beta 2, and beta 4 subunit genes, which they termed CCHNA1, CCHB2, and CCHB4, respectively, are located at different sites on proximal Chr 2, while the beta 3 subunit gene CCHB3 maps to Chr 15 near Wnt1. The authors suggested that these results, together with previous mapping data, indicated that the calcium channel genes are dispersed in the mouse genome, unlike the sodium channel genes, which are clustered.

The complete mRNA cds for murine voltage-dependent calcium channel beta-3 subunit (CCHB3) has been deposited in GenBank (GI/NID number: 836935; Accession number: U20372).

Nomura *et al.* (1998, *J. Biol. Chem.* 273:25967-25973) identified another channel protein, PKDL (also known as polycystic kidney disease 2-like 1; PKD2L1; PKD2L; polycystin L; polycystin-L), that encodes a member of the polycystin protein family. Nomura *et al.* (1998) found that PKDL had striking sequence and structural resemblance to the pore forming alpha-1 subunits of calcium channels. Chen *et al.* (1999, *Nature* 401:383-386) demonstrated that PKDL is a calcium-modulated nonselective cation channel that is permeable to sodium, potassium, and calcium ions. Channel activity was substantially increased when either the extracellular or intracellular calcium-ion concentration was raised, indicating that PKDL may act as a transducer of calcium-mediated

signaling *in vivo*. Chen *et al.* (1999) concluded that PKDL's large single-channel conductance and regulation by calcium ions distinguish it from other structurally related cation channels.

Nomura *et al.* (1998) found that the full-length PKDL transcript is expressed at high levels in fetal tissues, including kidney and liver, and is down-regulated in adult tissues. Nomura *et al.* (1998) mapped the PKDL gene to chromosome 10q24 by fluorescence *in situ* hybridization and linked it to D10S603 by radiation hybrid mapping. Wu *et al.* (1998, *Genomics* 54:564-568) found that PKDL is expressed in adult heart and skeletal muscle, brain, spleen, testis and retina. Wu *et al.* (1998) further mapped the PKDL gene to chromosome 10q25 by FISH and analysis of somatic cell hybrid mapping panel.

The complete mRNA cds for human polycystic kidney disease 2-like 1 (PKDL) has been deposited in GenBank (GI/NID number: 5305410; Accession number: AF073481).

Other channel proteins of interest include the glutamate receptor channels. Glutamate receptors mediate most excitatory neurotransmission in the brain, and molecular cloning studies have revealed several distinct families. Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system. Its physiologic action is exerted through the activation of ligand-gated ion channels and G protein-coupled membrane receptors.

Sakimura *et al.*, (1992, *Neuron* 8(2): 267-274), revealed by cloning and sequencing cDNA the presence and primary structure of a novel subunit of the mouse glutamate receptor (GLUR) channel, designated as gamma 2 (GLUR-gamma 2 was later referred to as the KA2 subunit). The gamma 2 subunit was found to have structural characteristics common to the neurotransmitter-gated ion channel family and to share a high amino acid sequence identity with the rat KA1 subunit, thus constituting the gamma subfamily of the glutamate receptor channel. Expression of the gamma 2 subunit together with the beta 2 subunit in Xenopus oocytes was found to yield functional glutamate receptor channels selective for kainate.

Glutamate-gated ionic channels are broadly classified into N-methyl-D-aspartate (NMDA) and non-NMDA types. Three related non-NMDA receptor subunit gene families have been defined: the AMPA-preferring family, e.g., Glutamate Receptor 1 (GLUR1), and the 2 kainate-preferring families. The Glutamate Receptors, Ionotropic Kainate types (GRIK1, GRIK2, and GRIK3) constitute one of the GRIK gene families, and KA1 and KA2 (also known as GRIK4 and GRIK5, respectively) constitute the second GRIK gene family. The KA1 and KA2 subunits display 68% identity in their amino acid sequence. They form functional heteromeric kainate-preferring ionic channels with the subunits encoded by the GRIK1, GRIK2, and GRIK3 genes but not with those falling in the AMP-preferring family. By Southern analysis of rat/mouse and human/mouse somatic cell hybrid panels and by fluorescence *in situ* hybridization, Szpirer *et al.*, (1994, *Proc. Nat. Acad. Sci.* 91:11849-11853), determined that the GRIK4 and GRIK5 genes are located on human 11q22.3 and human 19q13.2, respectively; on mouse chromosome 9 and mouse chromosome 7, respectively;

and on rat chromosome 8 and rat chromosome 1, respectively. The complete 3733 bp mRNA cds for the murine GRIK5 (*aka* KA2 or GLUR-gamma 2) gene has been deposited in GenBank (GI/NID number: 220415; Accession number: D10011).

Given the importance of channel proteins, a clear need exists for the elucidation of their functions, which information can be used in preventing, ameliorating or correcting dysfunctions or diseases associated therewith.

Summary of the Invention

The present invention generally relates to transgenic animals, as well as to compositions and methods relating to the characterization of gene function.

The present invention provides transgenic cells comprising a disruption in a channel protein target gene. The transgenic cells of the present invention are comprised of any cells capable of undergoing homologous recombination. Preferably, the cells of the present invention are stem cells and more preferably, embryonic stem (ES) cells, and most preferably, murine ES cells. According to one embodiment, the transgenic cells are produced by introducing a targeting construct into a stem cell to produce a homologous recombinant, resulting in a mutation of the target gene. In another embodiment, the transgenic cells are derived from the transgenic animals described below. The cells derived from the transgenic animals includes cells that are isolated or present in a tissue or organ, and any cell lines or any progeny thereof.

The present invention also provides a targeting construct and methods of producing the targeting construct that when introduced into stem cells produces a homologous recombinant. In one embodiment, the targeting construct of the present invention comprises first and second polynucleotide sequences that are homologous to the target gene. The targeting construct may also comprise a polynucleotide sequence that encodes a selectable marker that is preferably positioned between the two different homologous polynucleotide sequences in the construct. The targeting construct may also comprise other regulatory elements that can enhance homologous recombination.

The present invention further provides non-human transgenic animals and methods of producing such non-human transgenic animals comprising a disruption in a target gene. The transgenic animals of the present invention include transgenic animals that are heterozygous and homozygous for a null mutation in the target gene. In one aspect, the transgenic animals of the present invention are defective in the function of the target gene. In another aspect, the transgenic animals of the present invention comprise a phenotype associated with having a mutation in a target gene. Preferably, the transgenic animals are rodents and, most preferably, are mice.

The present invention also provides methods of identifying agents capable of affecting a phenotype of a transgenic animal. For example, a putative agent is administered to the transgenic animal and a response of the transgenic animal to the putative agent is measured and compared to the response of a "normal" or wild-type mouse, or alternatively compared to a transgenic animal

control (without agent administration). The invention further provides agents identified according to such methods. The present invention also provides methods of identifying agents useful as therapeutic agents for treating conditions associated with a disruption or other mutation (including naturally occurring mutations) of the target gene.

The present invention further provides a method of identifying agents having an effect on target expression or function. The method includes administering an effective amount of the agent to a transgenic animal, preferably a mouse. The method includes measuring a response of the transgenic animal, for example, to the agent, and comparing the response of the transgenic animal to a control animal, which may be, for example, a wild-type animal or alternatively, a transgenic animal control. Compounds that may have an effect on target expression or function may also be screened against cells in cell-based assays, for example, to identify such compounds.

The invention also provides cell lines comprising nucleic acid sequences of a target gene. Such cell lines may be capable of expressing such sequences by virtue of operable linkage to a promoter functional in the cell line. Preferably, expression of the target gene sequence is under the control of an inducible promoter. Also provided are methods of identifying agents that interact with the target gene, comprising the steps of contacting the target gene with an agent and detecting an agent/target gene complex. Such complexes can be detected by, for example, measuring expression of an operably linked detectable marker.

The invention further provides methods of treating diseases or conditions associated with a disruption in a target gene, and more particularly, to a disruption in the expression or function of the target gene. In a preferred embodiment, methods of the present invention involve treating diseases or conditions associated with a disruption in the target gene's expression or function, including administering to a subject in need, a therapeutic agent that affects target expression or function. In accordance with this embodiment, the method comprises administration of a therapeutically effective amount of a natural, synthetic, semi-synthetic, or recombinant target gene, target gene products or fragments thereof as well as natural, synthetic, semi-synthetic or recombinant analogs.

The present invention also provides compositions comprising or derived from ligands or other molecules or compounds that bind to or interact with target, including agonists or antagonists of target. Such agonists or antagonists of target include antibodies and antibody mimetics, as well as other molecules that can readily be identified by routine assays and experiments well known in the art.

The present invention further provides methods of treating diseases or conditions associated with disrupted targeted gene expression or function, wherein the methods comprise detecting and replacing through gene therapy mutated or otherwise defective or abnormal target genes.

Definitions

The term "gene" refers to (a) a gene containing at least one of the DNA sequences disclosed herein; (b) any DNA sequence that encodes the amino acid sequence encoded by the DNA sequences disclosed herein and/or; (c) any DNA sequence that hybridizes to the complement of the coding sequences disclosed herein. Preferably, the term includes coding as well as noncoding regions, and preferably includes all sequences necessary for normal gene expression including promoters, enhancers and other regulatory sequences.

The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleotides, ribonucleotides and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-, double-stranded and triple helical molecules.

"Oligonucleotide" refers to polynucleotides of between 5 and about 100 nucleotides of single- or double-stranded DNA. Oligonucleotides are also known as oligomers or oligos and may be isolated from genes, or chemically synthesized by methods known in the art. A "primer" refers to an oligonucleotide, usually single-stranded, that provides a 3'-hydroxyl end for the initiation of enzyme-mediated nucleic acid synthesis. The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentylnyluracil and 2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

A "fragment" of a polynucleotide is a polynucleotide comprised of at least 9 contiguous nucleotides, preferably at least 15 contiguous nucleotides and more preferably at least 45 nucleotides, of coding or non-coding sequences.

The term "gene targeting" refers to a type of homologous recombination that occurs when a fragment of genomic DNA is introduced into a mammalian cell and that fragment locates and recombines with endogenous homologous sequences.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of homologous nucleotide sequences.

The term "homologous" as used herein denotes a characteristic of a DNA sequence having at least about 70 percent sequence identity as compared to a reference sequence, typically at least about 85 percent sequence identity, preferably at least about 95 percent sequence identity, and more preferably about 98 percent sequence identity, and most preferably about 100 percent sequence identity as compared to a reference sequence. Homology can be determined using, for example, a "BLASTN" algorithm. It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome.

The term "target gene" (alternatively referred to as "target gene sequence" or "target DNA sequence" or "target sequence") refers to any nucleic acid molecule, polynucleotide, or gene to be modified by homologous recombination. The target sequence includes an intact gene, an exon or intron, a regulatory sequence or any region between genes. The target gene may comprise a portion of a particular gene or genetic locus in the individual's genomic DNA. As provided herein, "target gene" refers to a gene encoding a channel protein selected from the group consisting of: a Shab potassium channel-like gene; a CCHB3 voltage-dependant calcium channel beta-3 gene; a PKDL polycystin-L gene; or a glutamate receptor channel gamma-2 gene. A "Shab potassium channel-like gene" refers to a sequence comprising SEQ ID NO:1, or comprising the Shab potassium channel-like gene sequence identified in GenBank as Accession no.: AA718873; GI no.: 2731147, or a homolog or ortholog thereof. A "CCHB3 gene" refers to a sequence comprising SEQ ID NO:4, or comprising the CCHB3 gene sequence identified in GenBank as Accession no.: U20372; GI no.: 836935, or a homolog or ortholog thereof. A "PKDL gene" refers to a sequence comprising SEQ ID NO:8, or comprising the PKDL gene sequence identified in GenBank as Accession no.: AF073481; GI no.: 5305410, or a homolog or ortholog thereof. A "glutamate receptor channel gamma-2 gene" refers to a sequence comprising SEQ ID NO:12, or comprising the glutamate receptor channel gamma-2 gene sequence identified in GenBank as Accession no.: D10011; GI no.: 220415, or a homolog or ortholog thereof.

"Disruption" of a target gene occurs when a fragment of genomic DNA locates and recombines with an endogenous homologous sequence. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA sequence, or any combination thereof. Insertions include the insertion of entire genes, which may be of animal, plant, fungal, insect, prokaryotic, or viral origin. Disruption, for example, can alter or replace a promoter, enhancer, or splice site of a target gene, and can alter the normal gene product by inhibiting its production partially or completely or by enhancing the normal gene product's

activity. In a preferred embodiment, the disruption is a null disruption, wherein there is no significant expression of the target gene.

The term "construct" or "targeting construct" refers to an artificially assembled DNA segment to be transferred into a target tissue, cell line or animal, including human. Typically, the targeting construct will include a gene or a sequence of particular interest, a marker gene and appropriate control sequences. As provided herein, the targeting construct of the present invention comprises one of the following: a Shab potassium channel-like-specific targeting construct, a CCHB3-specific targeting construct, a PKDL-specific targeting construct, or a glutamate receptor channel gamma-2-specific targeting construct. A Shab potassium channel-like-specific targeting construct includes a DNA sequence homologous to at least one portion of a Shab potassium channel gene and is used to disrupt or knock out a Shab potassium channel gene in a host cell. A CCHB3-specific targeting construct includes a DNA sequence homologous to at least one portion of a CCHB3 gene and is used to disrupt or knock out a CCHB3 gene in a host cell. A PKDL-specific targeting construct includes a DNA sequence homologous to at least one portion of a PKDL gene and is used to disrupt or knock out a PKDL gene in a host cell. A glutamate receptor channel gamma-2-specific targeting construct includes a DNA sequence homologous to at least one portion of a glutamate receptor channel gamma-2 gene and is used to disrupt or knock out a glutamate receptor channel gamma-2 gene in a host cell.

The term "transgenic cell" refers to a cell containing within its genome a target gene that has been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting.

The term "transgenic animal" refers to an animal that contains within its genome a specific gene that has been disrupted or otherwise modified or mutated by the method of gene targeting. "Transgenic animal" includes both the heterozygous animal (*i.e.*, one defective allele and one wild-type allele) and the homozygous animal (*i.e.*, two defective alleles).

As used herein, the terms "selectable marker" and "positive selection marker" refer to a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. For example, plant and animal cells that express the introduced neomycin resistance (*Neo*^{r}) gene are resistant to the compound G418. Cells that do not carry the *Neo*^{r} gene marker are killed by G418. Other positive selection markers are known to, or are within the purview of, those of ordinary skill in the art.

A "host cell" includes an individual cell or cell culture that can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent due to natural, accidental, or deliberate mutation. A host cell includes cells transfected with the constructs of the present invention.

The term "modulates" as used herein refers to the decrease, inhibition, reduction, increase or enhancement of a target function, expression, activity, or alternatively a phenotype associated with a disruption in a target gene.

The term "ameliorates" refers to a decrease, reduction or elimination of a condition, disease, disorder, or phenotype, including an abnormality or symptom associated with a disruption in a target gene.

The term "abnormality" refers to any disease, disorder, condition, or phenotype in which a disruption of a target gene is implicated, including pathological conditions and behavioral observations.

Brief Description of the Drawings

Figure 1 shows the polynucleotide sequence for a murine EST (EST name: vt88g01.r1; GenBank Acc: AA718873; GenBank GI: 2731147) bearing sequence similarity and homology to Shab potassium ion channels (SEQ ID NO:1).

Figures 2 and 3 show the location and extent of the disrupted portion of the murine Shab potassium channel-like gene in certain of the transgenic mice of the present invention, as well as the nucleotide sequences flanking the *Neo*^{r} insert in the targeting construct. Specifically, Figure 3 shows the sequences identified as SEQ ID NO:2 and SEQ ID NO:3, which were used in the 5'- and 3'- targeting arms (which comprise homologous sequences) in the Shab potassium ion channel-like targeting construct, respectively.

Figure 4 shows the polynucleotide sequence for a murine CCHB3 voltage-dependant calcium channel beta-3 gene (SEQ ID NO:4).

Figure 5 shows the amino acid sequence for murine CCHB3 (SEQ ID NO:5).

Figures 6 and 7 show the location and extent of the disrupted portion of the murine CCHB3 gene in certain of the transgenic mice of the present invention, as well as the nucleotide sequences flanking the *Neo*^{r} insert in the targeting construct. Specifically, Figure 7 shows the sequences identified as SEQ ID NO:6 and SEQ ID NO:7, which were used in the 5'- and 3'- targeting arms (which comprise homologous sequences) in the CCHB3 targeting construct, respectively.

Figure 8 shows the polynucleotide sequence for a murine PKDL polycystin-L gene (SEQ ID NO:8).

Figure 9 shows the amino acid sequence for murine PKDL (SEQ ID NO:9).

Figures 10 and 11 show the location and extent of the disrupted portion of the murine PKDL gene in certain of the transgenic mice of the present invention, as well as the nucleotide sequences flanking the *Neo*^{r} insert in the targeting construct. Specifically, Figure 11 shows the sequences identified as SEQ ID NO:10 and SEQ ID NO:11, which were used in the 5'- and 3'- targeting arms (which comprise homologous sequences) in the PKDL targeting construct, respectively.

Figure 12 shows the polynucleotide sequence for a murine glutamate receptor channel gamma-2 gene (SEQ ID NO:12).

Figure 13 shows the amino acid sequence for murine glutamate receptor channel gamma-2 (SEQ ID NO:13).

Figures 14A-14B and 15 show the location and extent of the disrupted portion of the murine glutamate receptor channel gamma-2 gene in certain of the transgenic mice of the present invention, as well as the nucleotide sequences flanking the *Neo*^{r} insert in the targeting construct. Specifically, Figure 15 shows the sequences identified as SEQ ID NO:14 and SEQ ID NO:15, which were used in the 5'- and 3'- targeting arms (which comprise homologous sequences) in the glutamate receptor channel gamma-2 targeting construct, respectively.

Detailed Description of the Invention

The invention is based, in part, on the evaluation of the expression and role of genes and gene expression products, primarily those associated with a target gene. Among other uses or applications, the invention permits the definition of disease pathways and the identification of diagnostically and therapeutically useful targets. For example, genes that are mutated or down-regulated under disease conditions may be involved in causing or exacerbating the disease condition. Treatments directed at up-regulating the activity of such genes or treatments that involve alternate pathways, may ameliorate the disease condition.

Generation of Targeting Construct

The targeting construct of the present invention may be produced using standard methods known in the art. (*see, e.g.*, Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; E.N. Glover (eds.), 1985, *DNA Cloning: A Practical Approach*, Volumes I and II; M.J. Gait (ed.), 1984, *Oligonucleotide Synthesis*; B.D. Hames & S.J. Higgins (eds.), 1985, Nucleic Acid Hybridization; B.D. Hames & S.J. Higgins (eds.), 1984, *Transcription and Translation*; R.I. Freshney (ed.), 1986, Animal Cell Culture; Immobilized Cells and Enzymes, IRL Press, 1986; B. Perbal, 1984, A Practical Guide To Molecular Cloning; F.M. Ausubel *et al.*, 1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). For example, the targeting construct may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to *in vitro* mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like.

The targeting DNA can be constructed using techniques well known in the art. For example, the targeting DNA may be produced by chemical synthesis of oligonucleotides, nick-translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning

vectors harboring a sequence of interest (*e.g.*, a cloned cDNA or genomic DNA, synthetic DNA or from any of the aforementioned combination) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Moreover, the length of homology may be selected using known methods in the art. For example, selection may be based on the sequence composition and complexity of the predetermined endogenous target DNA sequence(s).

The targeting construct of the present invention typically comprises a first sequence homologous to a portion or region of the target gene and a second sequence homologous to a second portion or region of the target gene. The targeting construct may further comprise a positive selection marker, which is preferably positioned in between the first and the second DNA sequences that are homologous to a portion or region of the target DNA sequence. The positive selection marker may be operatively linked to a promoter and a polyadenylation signal.

Other regulatory sequences known in the art may be incorporated into the targeting construct to disrupt or control expression of a particular gene in a specific cell type. In addition, the targeting construct may also include a sequence coding for a screening marker, for example, green fluorescent protein (GFP), or another modified fluorescent protein.

Although the size of the homologous sequence is not critical and can range from as few as about 15-20 base pairs to as many as 100 kb, preferably each fragment is greater than about 1 kb in length, more preferably between about 1 and about 10 kb, and even more preferably between about 1 and about 5 kb. One of skill in the art will recognize that although larger fragments may increase the number of homologous recombination events in ES cells, larger fragments will also be more difficult to clone.

In a preferred embodiment of the present invention, the targeting construct is prepared directly from a plasmid genomic library using the methods described in pending U.S. Patent Application Ser. No.: 08/971,310, filed November 17, 1997, the disclosure of which is incorporated herein in its entirety. Generally, a sequence of interest is identified and isolated from a plasmid library in a single step using, for example, long-range PCR. Following isolation of this sequence, a second polynucleotide that will disrupt the target sequence can be readily inserted between two regions encoding the sequence of interest. In accordance with this aspect, the construct is generated in two steps by (1) amplifying (for example, using long-range PCR) sequences homologous to the target sequence, and (2) inserting another polynucleotide (for example a selectable marker) into the PCR product so that it is flanked by the homologous sequences. Typically, the vector is a plasmid from a plasmid genomic library. The completed construct is also typically a circular plasmid.

In another embodiment, the targeting construct is designed in accordance with the regulated positive selection method described in U.S. Patent Application Ser. No. 09/954,483, filed

September 17, 2001, the disclosure of which is incorporated herein in its entirety. The targeting construct is designed to include a PGK-*neo* fusion gene having two *lacO* sites, positioned in the PGK promoter and an NLS-*lacI* gene comprising a lac repressor fused to sequences encoding the NLS from the SV40 T antigen.

In another embodiment, the targeting construct may contain more than one selectable maker gene, including a negative selectable marker, such as the herpes simplex virus tk (HSV-tk) gene. The negative selectable marker may be operatively linked to a promoter and a polyadenylation signal. (*see, e.g.*, U.S. Patent No. 5,464,764; U.S. Patent No. 5,487,992; U.S. Patent No. 5,627,059; and U.S. Patent No. 5,631,153).

Generation of Cells and Confirmation of Homologous Recombination Events

Once an appropriate targeting construct has been prepared, the targeting construct may be introduced into an appropriate host cell using any method known in the art. Various techniques may be employed in the present invention, including, for example: pronuclear microinjection; retrovirus mediated gene transfer into germ lines; gene targeting in embryonic stem cells; electroporation of embryos; sperm-mediated gene transfer; and calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, bacterial protoplast fusion with intact cells, transfection, polycations, *e.g.*, polybrene, polyornithine, *etc.*, or the like (*see, e.g.*, U.S. Patent No. 4,873,191; Van der Putten, *et al.*, 1985, *Proc. Natl. Acad. Sci., USA* 82:6148-6152; Thompson, *et al.*, 1989, *Cell* 56:313-321; Lo, 1983, *Mol Cell. Biol.* 3:1803-1814; Lavitrano, *et al.*,1989, *Cell,* 57:717-723). Various techniques for transforming mammalian cells are known in the art. (*see, e.g.*, Gordon, 1989, *Intl. Rev. Cytol.*, 115:171-229; Keown *et al.*, 1989, *Methods in Enzymology*; Keown *et al.*, 1990, *Methods and Enzymology*, Vol. 185, pp. 527-537; Mansour *et al.*, 1988, *Nature*, 336:348-352).

In a preferred aspect of the present invention, the targeting construct is introduced into host cells by electroporation. In this process, electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the construct. The pores created during electroporation permit the uptake of macromolecules such as DNA. (*see, e.g.*, Potter, H., *et al.*, 1984, *Proc. Nat'l. Acad. Sci. U.S.A.* 81:7161-7165).

Any cell type capable of homologous recombination may be used in the practice of the present invention. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and ether eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants.

Preferred cell types include embryonic stem (ES) cells, which are typically obtained from pre-implantation embryos cultured *in vitro*. (*see, e.g.*, Evans, M. J., *et al.*, 1981, *Nature* 292:154-156; Bradley, M. O., *et al.*, 1984, *Nature* 309:255-258; Gossler *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:9065-9069; and Robertson, *et al.*, 1986, *Nature* 322:445-448). The ES cells are cultured and prepared for introduction of the targeting construct using methods well known to the skilled

artisan. (*see, e.g.*, Robertson, E. J. ed. "Teratocarcinomas and Embryonic Stem Cells, a Practical Approach", IRL Press, Washington D.C., 1987; Bradley *et al.*, 1986, *Current Topics in Devel. Biol.* 20:357-371; by Hogan *et al.*, in "Manipulating the Mouse Embryo": A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 1986; Thomas *et al.*, 1987, *Cell* 51:503; Koller *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:10730; Dorin *et al.*, 1992, *Transgenic Res.* 1:101; and Veis *et al.*, 1993, *Cell* 75:229). The ES cells that will be inserted with the targeting construct are derived from an embryo or blastocyst of the same species as the developing embryo into which they are to be introduced. ES cells are typically selected for their ability to integrate into the inner cell mass and contribute to the germ line of an individual when introduced into the mammal in an embryo at the blastocyst stage of development. Thus, any ES cell line having this capability is suitable for use in the practice of the present invention.

The present invention may also be used to knock out or otherwise modify or disrupt genes in other cell types, such as stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. These cells comprising a knock out, modification or disruption of a gene may be particularly useful in the study of target gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

After the targeting construct has been introduced into cells, the cells in which successful gene targeting has occurred are identified. Insertion of the targeting construct into the targeted gene is typically detected by identifying cells for expression of the marker gene. In a preferred embodiment, the cells transformed with the targeting construct of the present invention are subjected to treatment with an appropriate agent that selects against cells not expressing the selectable marker. Only those cells expressing the selectable marker gene survive and/or grow under certain conditions. For example, cells that express the introduced neomycin resistance gene are resistant to the compound G418, while cells that do not express the neo gene marker are killed by G418. If the targeting construct also comprises a screening marker such as GFP, homologous recombination can be identified through screening cell colonies under a fluorescent light. Cells that have undergone homologous recombination will have deleted the GFP gene and will not fluoresce.

If a regulated positive selection method is used in identifying homologous recombination events, the targeting construct is designed so that the expression of the selectable marker gene is regulated in a manner such that expression is inhibited following random integration but is permitted (derepressed) following homologous recombination. More particularly, the transfected cells are screened for expression of the *neo* gene, which requires that (1) the cell was successfully electroporated, and (2) *lac* repressor inhibition of *neo* transcription was relieved by homologous recombination. This method allows for the identification of transfected cells and homologous recombinants to occur in one step with the addition of a single drug.

Alternatively, a positive-negative selection technique may be used to select homologous recombinants. This technique involves a process in which a first drug is added to the cell population, for example, a neomycin-like drug to select for growth of transfected cells, *i.e.* positive selection. A second drug, such as FIAU is subsequently added to kill cells that express the negative selection marker, *i.e.* negative selection. Cells that contain and express the negative selection marker are killed by a selecting agent, whereas cells that do not contain and express the negative selection marker survive. For example, cells with non-homologous insertion of the construct express HSV thymidine kinase and therefore are sensitive to the herpes drugs such as gancyclovir (GANC) or FIAU (1-(2-deoxy 2-fluoro-B-D-arabinofluranosyl)-5-iodouracil). (*see, e.g.*, Mansour *et al.*, *Nature* 336:348-352: (1988); Capecchi, *Science* 244:1288-1292, (1989); Capecchi, *Trends in Genet.* 5:70-76 (1989)).

Successful recombination may be identified by analyzing the DNA of the selected cells to confirm homologous recombination. Various techniques known in the art, such as PCR and/or Southern analysis may be used to confirm homologous recombination events.

Homologous recombination may also be used to disrupt genes in stem cells, and other cell types, which are not totipotent embryonic stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. Such transgenic cells may be particularly useful in the study of target gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

In cells that are not totipotent, it may be desirable to knock out both copies of the target using methods that are known in the art. For example, cells comprising homologous recombination at a target locus that have been selected for expression of a positive selection marker (*e.g.*, *Neo*^{r}) and screened for non-random integration, can be further selected for multiple copies of the selectable marker gene by exposure to elevated levels of the selective agent (*e.g.*, G418). The cells are then analyzed for homozygosity at the target locus. Alternatively, a second construct can be generated with a different positive selection marker inserted between the two homologous sequences. The two constructs can be introduced into the cell either sequentially or simultaneously, followed by appropriate selection for each of the positive marker genes. The final cell is screened for homologous recombination of both alleles of the target.

Production of Transgenic Animals

Selected cells are then injected into a blastocyst (or other stage of development suitable for the purposes of creating a viable animal, such as, for example, a morula) of an animal (*e.g.*, a mouse) to form chimeras (*see e.g.*, Bradley, A. *in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed., IRL, Oxford, pp. 113-152 (1987)). Alternatively, selected ES cells can be allowed to aggregate with dissociated mouse embryo cells to form the aggregation

chimera. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Chimeric progeny harbouring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA. In one embodiment, chimeric progeny mice are used to generate a mouse with a heterozygous disruption in the target gene. Heterozygous transgenic mice can then be mated. It is well known in the art that typically ¼ of the offspring of such matings will have a homozygous disruption in the target gene.

The heterozygous and homozygous transgenic mice can then be compared to normal, wild-type mice to determine whether disruption of the target gene causes phenotypic changes, especially pathological changes. For example, heterozygous and homozygous mice may be evaluated for phenotypic changes by physical examination, necropsy, histology, clinical chemistry, complete blood count, body weight, organ weights, and cytological evaluation of bone marrow. Phenotypic changes may also comprise behavioral modifications or abnormalities.

In one embodiment, the phenotype (or phenotypic change) associated with a disruption in the target gene is placed into or stored in a database. Preferably, the database includes: (i) genotypic data (*e.g.*, identification of the disrupted gene) and (ii) phenotypic data (*e.g.*, phenotype(s) resulting from the gene disruption) associated with the genotypic data. The database is preferably electronic. In addition, the database is preferably combined with a search tool so that the database is searchable.

Conditional Transgenic Animals

The present invention further contemplates conditional transgenic or knockout animals, such as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes that cleave DNA at specific target sites (lox P sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable alternative site-specific recombinases have been described, and their genes can be used in accordance with the method of the present invention. Such recombinases include the Int recombinase of bacteriophage λ (with or without Xis) (Weisberg, R. *et al.*, in *Lambda II*, (Hendrix, R., *et al.*, Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-50 (1983), incorporated herein by reference); TpnI and the β-lactamase transposons (Mercier, *et al.*, *J. Bacteriol.*, 172:3745-57 (1990)); the Tn3 resolvase (Flanagan & Fennewald *J. Molec. Biol.*, 206:295-304 (1989); Stark, *et al.*, *Cell*, 58:779-90 (1989)); the yeast recombinases (Matsuzaki, *et al.*, *J. Bacteriol.*, 172:610-18 (1990)); the *B. subtilis* SpoIVC recombinase (Sato, *et al.*, *J. Bacteriol.* 172:1092-98 (1990)); the Flp recombinase (Schwartz & Sadowski, *J. Molec.Biol.*, 205:647-658 (1989); Parsons, *et al.*, *J. Biol. Chem.*, 265:4527-33 (1990); Golic & Lindquist, *Cell*, 59:499-509 (1989); Amin, *et al.*, *J. Molec. Biol.*, 214:55-72 (1990)); the Hin recombinase (Glasgow, *et al.*, *J. Biol. Chem.*, 264:10072-82 (1989)); immunoglobulin recombinases (Malynn, *et al.*, *Cell*,

54:453-460 (1988)); and the Cin recombinase (Haffter & Bickle, *EMBO J.*, 7:3991-3996 (1988); Hubner, *et al., J. Molec. Biol.*, 205:493-500 (1989)), all herein incorporated by reference. Such systems are discussed by Echols (*J. Biol. Chem.* 265:14697-14700 (1990)); de Villartay (*Nature,* 335:170-74 (1988)); Craig, (*Ann. Rev. Genet.*, 22:77-105 (1988)); Poyart-Salmeron, *et al.*, (*EMBO J.* 8:2425-33 (1989)); Hunger-Bertling, *et al.*,(*Mol Cell. Biochem.*, 92:107-16 (1990)); and Cregg & Madden (*Mol. Gen. Genet.*, 219:320-23 (1989)), all herein incorporated by reference.

Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski & Hess *J. Mol. Biol.* 259:1509-14 (1984), herein incorporated by reference). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/Du Pont. The cre gene (which encodes the Cre protein) has been cloned and expressed (Abremski, *et al., Cell* 32:1301-11 (1983), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg, *et al., Cold Spring Harbor Symp. Quant. Biol.* 45:297-309 (1981)), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess & Abremski *Proc. Natl. Acad. Sci. U.S.A.* 81:1026-29 (1984)). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski, *et al., Cell* 32:1301-11 (1983)). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt target genes, a fusion transcript can be produced when insertion of the positive selection marker occurs downstream (3') of the translation initiation site of the target gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

In one embodiment, purified recombinase enzyme is provided to the cell by direct microinjection. In another embodiment, recombinase is expressed from a co-transfected construct or vector in which the recombinase gene is operably linked to a functional promoter. An additional aspect of this embodiment is the use of tissue-specific or inducible recombinase constructs that allow the choice of when and where recombination occurs. One method for practicing the inducible forms of recombinase-mediated recombination involves the use of vectors that use inducible or tissue-specific promoters or other gene regulatory elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or activation of expression of the desired recombinase activity. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al., Proc. Natl. Acad. Sci. USA,* 93:3346-51 (1996); Furth, *et al., Proc. Natl. Acad. Sci. USA,* 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, promoters. Vectors incorporating such promoters would only express recombinase activity in cells that express the necessary transcription factors.

Models for Disease

The cell- and animal-based systems described herein can be utilized as models for diseases. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate disease animal models. In addition, cells from humans may be used. These systems may be used in a variety of applications. Such assays may be utilized as part of screening strategies designed to identify agents, such as compounds that are capable of ameliorating disease symptoms. Thus, the animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions that may be effective in treating disease.

Cell-based systems may be used to identify compounds that may act to ameliorate disease symptoms. For example, such cell systems may be exposed to a compound suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the disease cellular phenotypes has been altered to resemble a more normal or more wild-type, non-disease phenotype.

In addition, animal-based disease systems, such as those described herein, may be used to identify compounds capable of ameliorating disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions that may be effective in treating a disease or other phenotypic characteristic of the animal. For example, animal models may be exposed to a compound or agent suspected of exhibiting an ability to

ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with the disease. Exposure may involve treating mother animals during gestation of the model animals described herein, thereby exposing embryos or fetuses to the compound or agent that may prevent or ameliorate the disease or phenotype. Neonatal, juvenile, and adult animals can also be exposed.

More particularly, using the animal models of the invention, methods of identifying agents are provided, in which such agents can be identified on the basis of their ability to affect at least one phenotype associated with a disruption in a target gene. In one embodiment, the present invention provides a method of identifying agents having an effect on target expression or function. The method includes measuring a physiological response of the animal, for example, to the agent and comparing the physiological response of such animal to a control animal, wherein the physiological response of the animal comprising a disruption in a target as compared to the control animal indicates the specificity of the agent. A "physiological response" is any biological or physical parameter of an animal that can be measured. Molecular assays (*e.g.*, gene transcription, protein production and degradation rates), physical parameters (*e.g.*, exercise physiology tests, measurement of various parameters of respiration, measurement of heart rate or blood pressure and measurement of bleeding time), behavioral testing, and cellular assays (*e.g.*, immunohistochemical assays of cell surface markers, or the ability of cells to aggregate or proliferate) can be used to assess a physiological response.

The transgenic animals and cells of the present invention may be utilized as models for diseases, disorders, or conditions associated with phenotypes relating to a disruption in a target gene.

The present invention provides a unique animal model for testing and developing new treatments relating to the behavioral phenotypes. Analysis of the behavioral phenotype allows for the development of an animal model useful for testing, for instance, the efficacy of proposed genetic and pharmacological therapies for human genetic diseases, such as neurological, neuropsychological, or psychotic illnesses.

A statistical analysis of the various behaviors measured can be carried out using any conventional statistical program routinely used by those skilled in the art (such as, for example, "Analysis of Variance" or ANOVA). A "p" value of about 0.05 or less is generally considered to be statistically significant, although slightly higher p values may still be indicative of statistically significant differences. To statistically analyze abnormal behavior, a comparison is made between the behavior of a transgenic animal (or a group thereof) to the behavior of a wild-type mouse (or a group thereof), typically under certain prescribed conditions. "Abnormal behavior" as used herein refers to behavior exhibited by an animal having a disruption in the target gene, *e.g.* transgenic animal, which differs from an animal without a disruption in the target gene, *e.g.* wild-type mouse.

Abnormal behavior consists of any number of standard behaviors that can be objectively measured (or observed) and compared. In the case of comparison, it is preferred that the change be statistically significant to confirm that there is indeed a meaningful behavioral difference between the knockout animal and the wild-type control animal. Examples of behaviors that may be measured or observed include, but are not limited to, ataxia, rapid limb movement, eye movement, breathing, motor activity, cognition, emotional behaviors, social behaviors, hyperactivity, hypersensitivity, anxiety, impaired learning, abnormal reward behavior, and abnormal social interaction, such as aggression.

A series of tests may be used to measure the behavioral phenotype of the animal models of the present invention, including neurological and neuropsychological tests to identify abnormal behavior. These tests may be used to measure abnormal behavior relating to, for example, learning and memory, eating, pain, aggression, sexual reproduction, anxiety, depression, schizophrenia, and drug abuse. (*see, e.g.*, Crawley & Paylor, *Hormones and Behavior* 31:197-211 (1997)).

The social interaction test involves exposing a mouse to other animals in a variety of settings. The social behaviors of the animals (*e.g.*, touching, climbing, sniffing, and mating) are subsequently evaluated. Differences in behaviors can then be statistically analyzed and compared (*see, e.g.*, S. E. File, *et al.*, *Pharmacol. Bioch. Behav.* 22:941-944 (1985); R. R. Holson, *Phys. Behav.* 37:239-247 (1986)). Examplary behavioral tests include the following.

The mouse startle response test typically involves exposing the animal to a sensory (typically auditory) stimulus and measuring the startle response of the animal (*see, e.g.*, M. A. Geyer, *et al.*, *Brain Res. Bull.* 25:485-498 (1990); Paylor and Crawley, *Psychopharmacology* 132:169-180 (1997)). A pre-pulse inhibition test can also be used, in which the percent inhibition (from a normal startle response) is measured by "cueing" the animal first with a brief low-intensity pre-pulse prior to the startle pulse.

The electric shock test generally involves exposure to an electrified surface and measurement of subsequent behaviors such as, for example, motor activity, learning, social behaviors. The behaviors are measured and statistically analyzed using standard statistical tests. (*see, e.g.*, G. J. Kant, *et al., Pharm. Bioch. Behav.* 20:793-797 (1984); N. J. Leidenheimer, *et al.*, *Pharmacol. Bioch. Behav.* 30:351-355 (1988)).

The tail-pinch or immobilization test involves applying pressure to the tail of the animal and/or restraining the animal's movements. Motor activity, social behavior, and cognitive behavior are examples of the areas that are measured. (*see, e.g.*, M. Bertolucci D'Angic, *et al.*, *Neurochem.* 55:1208-1214 (1990)).

The novelty test generally comprises exposure to a novel environment and/or novel objects. The animal's motor behavior in the novel environment and/or around the novel object are measured and statistically analyzed. (*see, e.g.*, D. K. Reinstein, *et al.*, *Pharm. Bioch. Behav.* 17:193-202

(1982); B. Poucet, *Behav. Neurosci.* 103:1009-10016 (1989); R. R. Holson, *et al.*, *Phys. Behav.* 37:231-238 (1986)). This test may be used to detect visual processing deficiencies or defects.

The learned helplessness test involves exposure to stresses, for example, noxious stimuli, which cannot be affected by the animal's behavior. The animal's behavior can be statistically analyzed using various standard statistical tests. (*see, e.g.*, A. Leshner, *et al.*, *Behav. Neural Biol.* 26:497-501 (1979)).

Alternatively, a tail suspension test may be used, in which the "immobile" time of the mouse is measured when suspended "upside-down" by its tail. This is a measure of whether the animal struggles, an indicator of depression. In humans, depression is believed to result from feelings of a lack of control over one's life or situation. It is believed that a depressive state can be elicited in animals by repeatedly subjecting them to aversive situations over which they have no control. A condition of "learned helplessness" is eventually reached, in which the animal will stop trying to change its circumstances and simply accept its fate. Animals that stop struggling sooner are believed to be more prone to depression. Studies have shown that the administration of certain antidepressant drugs prior to testing increases the amount of time that animals struggle before giving up.

The Morris water-maze test comprises learning spatial orientations in water and subsequently measuring the animal's behaviors, such as, for example, by counting the number of incorrect choices. The behaviors measured are statistically analyzed using standard statistical tests. (*see, e.g.*, E. M. Spruijt, *et al.*, *Brain Res.* 527:192-197 (1990)).

Alternatively, a Y-shaped maze may be used (*see, e.g.*, McFarland, D.J., *Pharmacology, Biochemistry and Behavior* 32:723-726 (1989); Dellu, F., *et al.*, *Neurobiology of Learning and Memory* 73:31-48 (2000)). The Y-maze is generally believed to be a test of cognitive ability. The dimensions of each arm of the Y-maze can be, for example, approximately 40 cm x 8 cm x 20 cm, although other dimensions may be used. Each arm can also have, for example, sixteen equally spaced photobeams to automatically detect movement within the arms. At least two different tests can be performed using such a Y-maze. In a continuous Y-maze paradigm, mice are allowed to explore all three arms of a Y-maze for, *e.g.*, approximately 10 minutes. The animals are continuously tracked using photobeam detection grids, and the data can be used to measure spontaneous alteration and positive bias behavior. Spontaneous alteration refers to the natural tendency of a "normal" animal to visit the least familiar arm of a maze. An alternation is scored when the animal makes two consecutive turns in the same direction, thus representing a sequence of visits to the least recently entered arm of the maze. Position bias determines egocentrically defined responses by measuring the animal's tendency to favor turning in one direction over another. Therefore, the test can detect differences in an animal's ability to navigate on the basis of allocentric or egocentric mechanisms. The two-trial Y-maze memory test measures response to novelty and spatial memory

based on a free-choice exploration paradigm. During the first trial (acquisition), the animals are allowed to freely visit two arms of the Y-maze for, *e.g.*, approximately 15 minutes. The third arm is blocked off during this trial. The second trial (retrieval) is performed after an intertrial interval of, *e.g.*, approximately 2 hours. During the retrieval trial, the blocked arm is opened and the animal is allowed access to all three arms for, *e.g.*, approximately 5 minutes. Data are collected during the retrieval trial and analyzed for the number and duration of visits to each arm. Because the three arms of the maze are virtually identical, discrimination between novelty and familiarity is dependent on "environmental" spatial cues around the room relative to the position of each arm. Changes in arm entry and duration of time spent in the novel arm in a transgenic animal model may be indicative of a role of that gene in mediating novelty and recognition processes.

The passive avoidance or shuttle box test generally involves exposure to two or more environments, one of which is noxious, providing a choice to be learned by the animal. Behavioral measures include, for example, response latency, number of correct responses, and consistency of response. (*see, e.g.*, R. Ader, *et al.*, *Psychon. Sci.* 26:125-128 (1972); R. R. Holson, *Phys. Behav.* 37:221-230 (1986)). Alternatively, a zero-maze can be used. In a zero-maze, the animals can, for example, be placed in a closed quadrant of an elevated annular platform having, *e.g.*, 2 open and 2 closed quadrants, and are allowed to explore for approximately 5 minutes. This paradigm exploits an approach-avoidance conflict between normal exploratory activity and an aversion to open spaces in rodents. This test measures anxiety levels and can be used to evaluate the effectiveness of anti-anxiolytic drugs. The time spent in open quadrants versus closed quadrants may be recorded automatically, with, for example, the placement of photobeams at each transition site.

The food avoidance test involves exposure to novel food and objectively measuring, for example, food intake and intake latency. The behaviors measured are statistically analyzed using standard statistical tests. (*see, e.g.*, B. A. Campbell, *et al.*, *J. Comp. Physiol. Psychol.* 67:15-22 (1969)).

The elevated plus-maze test comprises exposure to a maze, without sides, on a platform, the animal's behavior is objectively measured by counting the number of maze entries and maze learning. The behavior is statistically analyzed using standard statistical tests. (*see, e.g.*, H. A. Baldwin, *et al.*, *Brain Res. Bull*, 20:603-606 (1988)).

The stimulant-induced hyperactivity test involves injection of stimulant drugs (*e.g.*, amphetamines, cocaine, PCP, and the like), and objectively measuring, for example, motor activity, social interactions, cognitive behavior. The animal's behaviors are statistically analyzed using standard statistical tests. (*see, e.g.*, P. B. S. Clarke, *et al.*, *Psychopharmacology* 96:511-520 (1988); P. Kuczenski, *et al.*, *J. Neuroscience* 11:2703-2712 (1991)).

The self-stimulation test generally comprises providing the mouse with the opportunity to regulate electrical and/or chemical stimuli to its own brain. Behavior is measured by frequency and

pattern of self-stimulation. Such behaviors are statistically analyzed using standard statistical tests. (*see, e.g.*, S. Nassif, *et al.*, *Brain Res.*, 332:247-257 (1985); W. L. Isaac, *et al.*, *Behav. Neurosci.* 103:345-355 (1989)).

The reward test involves shaping a variety of behaviors, *e.g.*, motor, cognitive, and social, measuring, for example, rapidity and reliability of behavioral change, and statistically analyzing the behaviors measured. (*see, e.g.*, L. E. Jarrard, *et al.*, Exp. Brain Res. 61:519-530 (1986)).

The DRL (differential reinforcement to low rates of responding) performance test involves exposure to intermittent reward paradigms and measuring the number of proper responses, *e.g.*, lever pressing. Such behavior is statistically analyzed using standard statistical tests. (*see, e.g.*, J. D. Sinden, *et al.*, Behav. Neurosci. 100:320-329 (1986); V. Nalwa, *et al.*, Behav Brain Res. 17:73-76 (1985); and A. J. Nonneman, *et al.*, J. Comp. Physiol. Psych. 95:588-602 (1981)).

The spatial learning test involves exposure to a complex novel environment, measuring the rapidity and extent of spatial learning, and statistically analyzing the behaviors measured. (*see, e.g.*, N. Pitsikas, *et al.*, Pharm. Bioch. Behav. 38:931-934 (1991); B. poucet, *et al.*, Brain Res. 37:269-280 (1990); D. Christie, *et al.*, Brain Res. 37:263-268 (1990); and F. Van Haaren, *et al.*, Behav. Neurosci. 102:481-488 (1988)). Alternatively, an open-field (of) test may be used, in which the greater distance traveled for a given amount of time is a measure of the activity level and anxiety of the animal. When the open field is a novel environment, it is believed that an approach-avoidance situation is created, in which the animal is "torn" between the drive to explore and the drive to protect itself. Because the chamber is lighted and has no places to hide other than the corners, it is expected that a "normal" mouse will spend more time in the corners and around the periphery than it will in the center where there is no place to hide. "Normal" mice will, however, venture into the central regions as they explore more and more of the chamber. It can then be extrapolated that especially anxious mice will spend most of their time in the corners, with relatively little or no exploration of the central region, whereas bold (*i.e.*, less anxious) mice will travel a greater distance, showing little preference for the periphery versus the central region.

The visual, somatosensory and auditory neglect tests generally comprise exposure to a sensory stimulus, objectively measuring, for example, orientating responses, and statistically analyzing the behaviors measured. (*see, e.g.*, J. M. Vargo, *et al.*, Exp. Neurol. 102:199-209 (1988)).

The consummatory behavior test generally comprises feeding and drinking, and objectively measuring quantity of consumption. The behavior measured is statistically analyzed using standard statistical tests. (*see, e.g.*, P. J. Fletcher, *et al.*, Psychopharmacol. 102:301-308 (1990); M. G. Corda, *et al.*,, Proc. Nat'l Acad. Sci. USA 80:2072-2076 (1983)).

A visual discrimination test can also be used to evaluate the visual processing of an animal. One or two similar objects are placed in an open field and the animal is allowed to explore for about 5-10 minutes. The time spent exploring each object (proximity to, *i.e.*, movement within, *e.g.*, about

3-5 cm of the object is considered exploration of an object) is recorded. The animal is then removed from the open field, and the objects are replaced by a similar object and a novel object. The animal is returned to the open field and the percent time spent exploring the novel object over the old object is measured (again, over about a 5-10 minute span). "Normal" animals will typically spend a higher percentage of time exploring the novel object rather than the old object. If a delay is imposed between sampling and testing, the memory task becomes more hippocampal-dependent. If no delay is imposed, the task is more based on simple visual discrimination. This test can also be used for olfactory discrimination, in which the objects (preferably, simple blocks) can be sprayed or otherwise treated to hold an odor. This test can also be used to determine if the animal can make gustatory discriminations; animals that return to the previously eaten food instead of novel food exhibit gustatory neophobia.

A hot plate analgesia test can be used to evaluate an animal's sensitivity to heat or painful stimuli. For example, a mouse can be placed on an approximately 55°C hot plate and the mouse's response latency (*e.g.*, time to pick up and lick a hind paw) can be recorded. These responses are not reflexes, but rather "higher" responses requiring cortical involvement. This test may be used to evaluate a nociceptive disorder.

An accelerating rotarod test may be used to measure coordination and balance in mice. Animals can be, for example, placed on a rod that acts like a rotating treadmill (or rolling log). The rotarod can be made to rotate slowly at first and then progressively faster until it reaches a speed of, *e.g.*, approximately 60 rpm. The mice must continually reposition themselves in order to avoid falling off. The animals are preferably tested in at least three trials, a minimum of 20 minutes apart. Those mice that are able to stay on the rod the longest are believed to have better coordination and balance.

A metrazol administration test can be used to screen animals for varying susceptibilities to seizures or similar events. For example, a 5mg/ml solution of metrazol can be infused through the tail vein of a mouse at a rate of, *e.g.*, approximately 0.375 ml/min. The infusion will cause all mice to experience seizures, followed by death. Those mice that enter the seizure stage the soonest are believed to be more prone to seizures. Four distinct physiological stages can be recorded: soon after the start of infusion, the mice will exhibit a noticeable "twitch", followed by a series of seizures, ending in a final tensing of the body known as "tonic extension", which is followed by death.

Target Gene Products

The present invention further contemplates use of the target gene sequence to produce target gene products. target gene products may include proteins that represent functionally equivalent gene products. Such an equivalent gene product may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the gene sequences described herein, but which result in a silent change, thus producing a functionally equivalent target gene

product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar *in vivo* activity as the endogenous gene products encoded by the target gene sequences. Alternatively, when utilized as part of an assay, "functionally equivalent" may refer to peptides capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the endogenous gene product would.

Other protein products useful according to the methods of the invention are peptides derived from or based on the target gene produced by recombinant or synthetic means (derived peptides).

Target gene products may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the gene polypeptides and peptides of the invention by expressing nucleic acid encoding gene sequences are described herein. Methods that are well known to those skilled in the art can be used to construct expression vectors containing gene protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination (*see, e.g.*, Sambrook, *et al.*, 1989, *supra*, and Ausubel, *et al.*, 1989, supra). Alternatively, RNA capable of encoding gene protein sequences may be chemically synthesized using, for example, automated synthesizers (*see, e.g.* Oligonucleotide Synthesis: A Practical Approach, Gait, M. J. ed., IRL Press, Oxford (1984)).

A variety of host-expression vector systems may be utilized to express the gene coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells that may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the gene protein of the invention *in situ*. These include but are not limited to microorganisms such as bacteria (*e.g., E. coli, B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing gene protein coding sequences; yeast (*e.g. Saccharomyces, Pichia*) transformed with recombinant yeast expression vectors containing the gene protein coding sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the gene protein coding sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing

gene protein coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionine promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, *EMBO J.*, 2:1791-94 (1983)), in which the gene protein coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, *Nucleic Acids Res.*, 13:3101-09 (1985); Van Heeke *et al.*, *J. Biol. Chem.*, 264:5503-9 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a preferred embodiment, full length cDNA sequences are appended with in-frame Bam HI sites at the amino terminus and Eco RI sites at the carboxyl terminus using standard PCR methodologies (Innis, *et al.* (eds) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego (1990)) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radio-active labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.*, *EMBO J.*, 4: 1075-80 (1985); Zabeau *et al.*, *EMBO J.*, 1: 1217-24 (1982)).

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in Spodoptera frugiperda cells. The gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (*see, e.g.*, Smith, *et al.*, *J. Virol.* 46: 584-93 (1983); U.S. Patent No. 4,745,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gene protein in infected hosts. (*e.g.*, *see* Logan *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3655-59 (1984)). Specific initiation signals may also be required for efficient translation of inserted gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (*see* Bitter, *et al.*, *Methods in Enzymol.*, 153:516-44 (1987)).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the gene protein may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells that stably integrate the

plasmid into their chromosomes and grow, to form foci, which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene protein.

In a preferred embodiment, timing and/or quantity of expression of the recombinant protein can be controlled using an inducible expression construct. Inducible constructs and systems for inducible expression of recombinant proteins will be well known to those skilled in the art. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. In one in embodiment, a Tet inducible gene expression system is utilized. (Gossen *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5547-51 (1992); Gossen, *et al.*, *Science*, 268:1766-69 (1995)). Tet Expression Systems are based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon—the tetracycline repressor protein (TetR) and the tetracycline operator sequence (*tetO*) to which TetR binds. Using such a system, expression of the recombinant protein is placed under the control of the *tetO* operator sequence and transfected or transformed into a host cell. In the presence of TetR, which is co-transfected into the host cell, expression of the recombinant protein is repressed due to binding of the TetR protein to the *tetO* regulatory element. High-level, regulated gene expression can then be induced in response to varying concentrations of tetracycline (Tc) or Tc derivatives such as doxycycline (Dox), which compete with *tetO* elements for binding to TetR. Constructs and materials for tet inducible gene expression are available commercially from CLONTECH Laboratories, Inc., Palo Alto, CA.

When used as a component in an assay system, the gene protein may be labeled, either directly or indirectly, to facilitate detection of a complex formed between the gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ^{125}I; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels. Where recombinant DNA technology is used to produce the gene protein for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the gene product. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

Production of Antibodies

Described herein are methods for the production of antibodies capable of specifically recognizing one or more epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used, for example, in the detection of a target gene in a biological sample, or, alternatively, as a method for the inhibition of abnormal target gene activity. Thus, such antibodies may be utilized as part of disease treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of target gene proteins, or for the presence of abnormal forms of such proteins.

For the production of antibodies, various host animals may be immunized by injection with the target gene, its expression product or a portion thereof. Such host animals may include but are not limited to rabbits, mice, rats, goats and chickens, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum.*

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Köhler and Milstein, *Nature*, 256:495-7 (1975); and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor, *et al.*, *Immunology Today*, 4:72 (1983); Cote, *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:2026-30 (1983)), and the EBV-hybridoma technique (Cole, *et al.*, in Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., New York, pp. 77-96 (1985)). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison, *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Takeda, *et al.*, *Nature*, 314:452-54 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with

genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, *Science* 242:423-26 (1988); Huston, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:5879-83 (1988); and Ward, *et al.*, *Nature*, 334:544-46 (1989)) can be adapted to produce gene-single chain antibodies. Single chain antibodies are typically formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragments that can be produced by pepsin digestion of the antibody molecule and the Fab fragments that can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse, *et al.*, *Science*, 246:1275-81 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Screening Methods

The present invention may be employed in a process for screening for agents such as agonists, *i.e.*, agents that bind to and activate target polypeptides, or antagonists, *i.e.*, inhibit the activity or interaction of target polypeptides with its ligand. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures as known in the art. Any methods routinely used to identify and screen for agents that can modulate receptors may be used in accordance with the present invention.

The present invention provides methods for identifying and screening for agents that modulate target expression or function. More particularly, cells that contain and express target gene sequences may be used to screen for therapeutic agents. Such cells may include non-recombinant monocyte cell lines, such as U937 (ATCC# CRL-1593), THP-1 (ATCC# TIB-202), and P388D1 (ATCC# TIB-63); endothelial cells such as HUVEC's and bovine aortic endothelial cells (BAEC's); as well as generic mammalian cell lines such as HeLa cells and COS cells, *e.g.*, COS-7 (ATCC# CRL-1651). Further, such cells may include recombinant, transgenic cell lines. For example, the transgenic mice of the invention may be used to generate cell lines, containing one or more cell types involved in a disease, that can be used as cell culture models for that disorder. While cells, tissues, and primary cultures derived from the disease transgenic animals of the invention may be utilized, the generation of continuous cell lines is preferred. For examples of techniques that may be used to derive a continuous cell line from the transgenic animals, see Small, *et al.*, *Mol. Cell Biol.*, 5:642-48 (1985).

Target gene sequences may be introduced into and overexpressed in, the genome of the cell of interest. In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence that is capable of driving gene expression in the cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation. target gene sequences may also be disrupted or underexpressed. Cells having target gene disruptions or underexpressed target gene sequences may be used, for example, to screen for agents capable of affecting alternative pathways that compensate for any loss of function attributable to the disruption or underexpression.

In vitro systems may be designed to identify compounds capable of binding the target gene products. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; (*see e.g.*, Lam, *et al.*, *Nature*, 354:82-4 (1991)), phosphopeptides (in, for example, the form of random or partially degenerate, directed phosphopeptide libraries; *see*, *e.g.*, Songyang, *et al.*, *Cell*, 72:767-78 (1993)), antibodies, and small organic or inorganic molecules. Compounds identified may be useful, for example, in modulating the activity of target gene proteins, preferably mutant target gene proteins; elaborating the biological function of the target gene protein; or screening for compounds that disrupt normal target gene interactions or themselves disrupt such interactions.

The principle of the assays used to identify compounds that bind to the target gene protein involves preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the target gene protein or the test substance onto a solid phase and detecting target protein/test substance complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the target gene protein may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

In practice, microtitre plates are conveniently utilized. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-

labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific for target gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

Compounds that are shown to bind to a particular target gene product through one of the methods described above can be further tested for their ability to elicit a biochemical response from the target gene protein. Agonists, antagonists and/or inhibitors of the expression product can be identified utilizing assays well known in the art.

Antisense, Ribozymes, and Antibodies

Other agents that may be used as therapeutics include the target gene, its expression product(s) and functional fragments thereof. Additionally, agents that reduce or inhibit mutant target gene activity may be used to ameliorate disease symptoms. Such agents include antisense, ribozyme, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, *e.g.*, between the -10 and +10 regions of the target gene nucleotide sequence of interest, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to the target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Patent No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding target gene proteins.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites that include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides

corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription should be single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

It is possible that the antisense, ribozyme, and/or triple helix molecules described herein may reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by both normal and mutant target gene alleles. In order to ensure that substantially normal levels of target gene activity are maintained, nucleic acid molecules that encode and express target polypeptides exhibiting normal activity may be introduced into cells that do not contain sequences susceptible to whatever antisense, ribozyme, or triple helix treatments are being utilized. Alternatively, it may be preferable to coadminister normal target protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors that incorporate suitable RNA polymerase promoters such as the T7 or SP6

polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various well-known modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Antibodies that are both specific for target protein, and in particular, the mutant target protein, and interfere with its activity may be used to inhibit mutant target gene function. Such antibodies may be generated against the proteins themselves or against peptides corresponding to portions of the proteins using standard techniques known in the art and as also described herein. Such antibodies include but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, antibody mimetics, etc.

In instances where the target protein is intracellular and whole antibodies are used, internalizing antibodies may be preferred. However, lipofectin liposomes may be used to deliver the antibody or a fragment of the Fab region that binds to the target gene epitope into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target or expanded target protein's binding domain is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of the antibody that binds to the target protein may be used. Such peptides may be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (*see, e.g.*, Creighton, Proteins: Structures and Molecular Principles (1984) W.H. Freeman, New York 1983, *supra*; and Sambrook, *et al.*, 1989, supra). Alternatively, single chain neutralizing antibodies that bind to intracellular target gene epitopes may also be administered. Such single chain antibodies may be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques such as those described in Marasco, *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:7889-93 (1993).

RNA sequences encoding target protein may be directly administered to a patient exhibiting disease symptoms, at a concentration sufficient to produce a level of target protein such that disease symptoms are ameliorated. Patients may be treated by gene replacement therapy. One or more copies of a normal target gene, or a portion of the gene that directs the production of a normal target protein with target gene function, may be inserted into cells using vectors that include, but are not limited to adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes. Additionally, techniques such as those described above may be utilized for the introduction of normal target gene sequences into human cells.

Cells, preferably autologous cells, containing normal target gene expressing gene sequences may then be introduced or reintroduced into the patient at positions that allow for the amelioration of disease symptoms.

Pharmaceutical Compositions, Effective Dosages, and Routes of Administration

The identified compounds that inhibit target mutant gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to treat or ameliorate the disease. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disease.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, topical, subcutaneous, intraperitoneal, intraveneous, intrapleural, intraoccular, intraarterial, or rectal administration. It is also contemplated that pharmaceutical compositions may be administered with other products that potentiate the activity of the compound and optionally, may include other therapeutic ingredients.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients

such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of *e.g.* gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. Oral ingestion is possibly the easiest method of taking any medication. Such a route of administration, is generally simple and straightforward and is frequently the least inconvenient or unpleasant route of administration from the patient's point of view. However, this involves passing the material through the stomach, which is a hostile environment for many materials, including

proteins and other biologically active compositions. As the acidic, hydrolytic and proteolytic environment of the stomach has evolved efficiently to digest proteinaceous materials into amino acids and oligopeptides for subsequent anabolism, it is hardly surprising that very little or any of a wide variety of biologically active proteinaceous material, if simply taken orally, would survive its passage through the stomach to be taken up by the body in the small intestine. The result, is that many proteinaceous medicaments must be taken in through another method, such as parenterally, often by subcutaneous, intramuscular or intravenous injection.

Pharmaceutical compositions may also include various buffers (*e.g.*, Tris, acetate, phosphate), solubilizers (*e.g.*, Tween, Polysorbate), carriers such as human serum albumin, preservatives (thimerosol, benzyl alcohol) and anti-oxidants such as ascorbic acid in order to stabilize pharmaceutical activity. The stabilizing agent may be a detergent, such as tween-20, tween-80, NP-40 or Triton X-100. EBP may also be incorporated into particulate preparations of polymeric compounds for controlled delivery to a patient over an extended period of time. A more extensive survey of components in pharmaceutical compositions is found in Remington's Pharmaceutical Sciences, 18th ed., A. R. Gennaro, ed., Mack Publishing, Easton, Pa. (1990).

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example, subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device that may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

Diagnostics

A variety of methods may be employed to diagnose disease conditions associated with the target gene. Specifically, reagents may be used, for example, for the detection of the presence of target gene mutations, or the detection of either over- or under- expression of target gene mRNA.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type target gene locus is detected. In addition, the method can be performed by detecting the wild-type target gene locus and confirming the lack of a predisposition or neoplasia. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those that occur only in certain tissues, *e.g.*, in tumor tissue,

and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state may be indicated. However, if both alleles are mutated, then a late neoplastic state may be indicated. The finding of gene mutations thus provides both diagnostic and prognostic information. a target gene allele that is not deleted (*e.g.*, that found on the sister chromosome to a chromosome carrying a target gene deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. Mutations found in tumor tissues may be linked to decreased expression of the target gene product. However, mutations leading to non-functional gene products may also be linked to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the target gene product, or a decrease in mRNA stability or translation efficiency.

One test available for detecting mutations in a candidate locus is to directly compare genomic target sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, *e.g.*, by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene. Mutations from cancer patients falling outside the coding region of the target gene can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the target gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific gene nucleic acid or anti-gene antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings, to diagnose patients exhibiting disease symptoms or at risk for developing disease.

Any cell type or tissue, including brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord, eye, Harderian gland, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary, uterus and white fat, in which the gene is expressed may be utilized in the diagnostics described below.

DNA or RNA from the cell type or tissue to be analyzed may easily be isolated using procedures that are well known to those in the art. Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as

probes and/or primers for such *in situ* procedures (*see*, for example, Nuovo, PCR *In Situ* Hybridization: Protocols and Applications, Raven Press, N.Y. (1992)).

Gene nucleotide sequences, either RNA or DNA, may, for example, be used in hybridization or amplification assays of biological samples to detect disease-related gene structures and expression. Such assays may include, but are not limited to, Southern or Northern analyses, restriction fragment length polymorphism assays, single stranded conformational polymorphism analyses, *in situ* hybridization assays, and polymerase chain reaction analyses. Such analyses may reveal both quantitative aspects of the expression pattern of the gene, and qualitative aspects of the gene expression and/or gene composition. That is, such aspects may include, for example, point mutations, insertions, deletions, chromosomal rearrangements, and/or activation or inactivation of gene expression.

Preferred diagnostic methods for the detection of gene-specific nucleic acid molecules may involve for example, contacting and incubating nucleic acids, derived from the cell type or tissue being analyzed, with one or more labeled nucleic acid reagents under conditions favorable for the specific annealing of these reagents to their complementary sequences within the nucleic acid molecule of interest. Preferably, the lengths of these nucleic acid reagents are at least 9 to 30 nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid:fingerprint molecule hybrid. The presence of nucleic acids from the fingerprint tissue that have hybridized, if any such molecules exist, is then detected. Using such a detection scheme, the nucleic acid from the tissue or cell type of interest may be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtitre plate or polystyrene beads. In this case, after incubation, non-annealed, labeled nucleic acid reagents are easily removed. Detection of the remaining, annealed, labeled nucleic acid reagents is accomplished using standard techniques well-known to those in the art.

Alternative diagnostic methods for the detection of gene-specific nucleic acid molecules may involve their amplification, *e.g.*, by PCR (the experimental embodiment set forth in Mullis U.S. Patent No. 4,683,202 (1987)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-93 (1991)), self sustained sequence replication (Guatelli, *et al.*, *Proc. Natl. Acad. Sci.* USA, 87:1874-78 (1990)), transcriptional amplification system (Kwoh, *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173-77 (1989)), Q-Beta Replicase (Lizardi *et al.*, *Bio/Technology*, 6:1197 (1988)), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment of such a detection scheme, a cDNA molecule is obtained from an RNA molecule of interest (*e.g.*, by reverse transcription of the RNA molecule into cDNA). Cell types or tissues from which such RNA may be isolated include any tissue in which wild-type fingerprint

gene is known to be expressed, including, but not limited, to brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord, eye, Harderian gland, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary, uterus and white fat. A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method may be chosen from among the gene nucleic acid reagents described herein. The preferred lengths of such nucleic acid reagents are at least 15-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

Antibodies directed against wild-type or mutant gene peptides may also be used as disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of gene protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of fingerprint gene protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant fingerprint gene protein relative to the normal fingerprint gene protein.

Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques that are well known to those of skill in the art, including but not limited to western blot analysis. For a detailed explanation of methods for carrying out western blot analysis, see Sambrook, *et al.* (1989) *supra*, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

Preferred diagnostic methods for the detection of wild-type or mutant gene peptide molecules may involve, for example, immunoassays wherein fingerprint gene peptides are detected by their interaction with an anti-fingerprint gene-specific peptide antibody.

For example, antibodies, or fragments of antibodies useful in the present invention may be used to quantitatively or qualitatively detect the presence of wild-type or mutant gene peptides. This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. Such techniques are especially preferred if the fingerprint gene peptides are expressed on the cell surface.

The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of fingerprint gene peptides. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the fingerprint gene peptides, but also their distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Immunoassays for wild-type, mutant, or expanded fingerprint gene peptides typically comprise incubating a biological sample, such as a biological fluid, a tissue extract, freshly harvested cells, or cells that have been incubated in tissue culture, in the presence of a detectably labeled antibody capable of identifying fingerprint gene peptides, and detecting the bound antibody by any of a number of techniques well known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene-specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

The terms "solid phase support or carrier" are intended to encompass any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

The binding activity of a given lot of anti-wild-type or -mutant fingerprint gene peptide antibody may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

One of the ways in which the gene peptide-specific antibody can be detectably labeled is by linking the same to an enzyme and using it in an enzyme immunoassay (EIA) (Voller, *Ric Clin Lab*, 8:289-98 (1978) ["The Enzyme Linked Immunosorbent Assay (ELISA)", Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, Md.]; Voller, *et al.*, *J. Clin. Pathol.*, 31:507-20 (1978); Butler, *Meth. Enzymol.*, 73:482-523 (1981); Maggio (ed.), Enzyme Immunoassay, CRC Press, Boca Raton, Fla. (1980); Ishikawa, *et al.*, (eds.) Enzyme Immunoassay, Igaku-Shoin, Tokyo (1981)). The enzyme that is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety that can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes that can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild-type, mutant, or expanded peptides through the use of a radioimmunoassay (RIA) (*see, e.g.*, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediamine-tetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of

particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Throughout this application, various publications, patents and published patent applications are referred to by an identifying citation. The disclosures of these publications, patents and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

The following examples are intended only to illustrate the present invention and should in no way be construed as limiting the subject invention.

Examples

Example 1: Generation and Analysis of Mice Comprising Shab Potassium Ion Channel-like Gene Disruptions

To investigate the role of Shab potassium ion channels, disruptions in their target genes were produced by homologous recombination, and transgenic mice comprising such disruptions were created. More particularly, as shown in Figure 3, a target-specific construct having the ability to disrupt this target gene comprising SEQ ID NO:1 was created using in the targeting arms of the construct the oligonucleotide sequences identified herein as SEQ ID NO:2 and SEQ ID NO:3.

The targeting construct was introduced into ES cells derived from the 129/OlaHsd mouse substrain. F1 mice were generated by breeding with C57BL/6 females. F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

The transgenic mice comprising disruptions in target genes were analyzed for phenotypic changes and expression patterns.

RT-PCR Expression. Total RNA was isolated from the organs or tissues from adult C57BL/6 wild-type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers.

Low levels of RNA transcripts were detectable in brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord, eye, gallbladder, urinary bladder, pituitary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovaries, uterus and white fat. No RNA transcripts were

detectable in Harderian glands, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, adrenal gland and salivary gland.

LacZ Reporter Gene Expression. In general, tissues from 7-12 week old heterozygous mutant mice were analyzed for lacZ expression. Organs from heterozygous mutant mice were frozen, sectioned (10 μm), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase, followed by a Nuclear Fast Red counterstaining.

In addition, for brain, wholemount staining was performed. The dissected brain was cut longitudinally, fixed and stained using X-Gal as the substrate for beta-galactosidase. The reaction was stopped by washing the brain in PBS and then fixed in PBS-buffered formaldehyde.

Wild-type control tissues were also stained for lacZ expression to reveal any background or signals due to endogenous beta-galactosidase activity. The following tissues can show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

LacZ (beta-galactosidase) expression was detectable in brain and spinal cord. LacZ expression was not detected in: sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gall bladder, pancreas, kidney, urinary bladder, esophagus, trachea, larynx, thyroid gland, pituitary gland, adrenal glands, salivary glands, tongue, skeletal muscle, skin, male and female reproductive systems.

Brain

In wholemount staining strong lacZ signals were detectable in the habenular nuclei and in a region which may represent the red nucleus magnocellular. Weaker staining was observed in cerebellum, along the anterior cerebellar lobes and in brainstem. In frozen sections of brain, very strong lacZ staining was detectable in the medial habenular nucleus.

Spinal Cord

Strong X-Gal staining was detectable in specific motor neurons.

Example 2: Generation and Analysis of Mice Comprising CCHB3 Gene Disruptions

To investigate the role of CCHB3 genes, disruptions in their target genes were produced by homologous recombination, and transgenic mice comprising such disruptions were created. More particularly, as shown in Figure 7, a target-specific construct having the ability to disrupt this target gene comprising SEQ ID NO:4 was created using in the targeting arms of the construct the oligonucleotide sequences identified herein as SEQ ID NO:6 and SEQ ID NO:7.

The targeting construct was introduced into ES cells derived from the 129/ mouse substrain. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

The transgenic mice comprising disruptions in CCHB3 genes were analyzed for phenotypic changes and expression patterns.

RT-PCR Expression. Total RNA was isolated from the organs or tissues from adult C57BL/6 wild-type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers.

RNA transcripts were detectable in brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, eye, heart, lung, kidneys, thymus, lymph nodes, bone marrow, skin, gall bladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus. RNA transcripts were not detectable in liver, pancreas, spleen and epididymis.

LacZ Reporter Gene Expression. In general, tissues from 7-12 week old heterozygous mutant mice were analyzed for lacZ expression. Organs from heterozygous mutant mice were frozen, sectioned (10 μm), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase, followed by a Nuclear Fast Red counterstaining.

In addition, for brain, wholemount staining was performed. The dissected brain was cut longitudinally, fixed and stained using X-Gal as the substrate for beta-galactosidase. The reaction was stopped by washing the brain in PBS and then fixed in PBS-buffered formaldehyde.

Wild-type control tissues were also stained for lacZ expression to reveal any background or signals due to endogenous beta-galactosidase activity. The following tissues can show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

LacZ (beta-galactosidase) expression was detectable in brain. LacZ expression was not detected in: spinal cord, sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gall bladder, pancreas, kidney, urinary bladder, larynx, trachea, esophagus, pituitary gland, thyroid gland, parathyroid gland, adrenal glands, salivary glands, tongue, skeletal muscle, skin, male and female reproductive systems.

Brain

In wholemount staining, strong X-Gal signals were detectable in choroid plexus. Weak X-Gal signals were detectable in olfactory bulbs, cortex and cerebellum. On frozen sections faint lacZ expression was detectable in caudate putamen.

Example 3: Generation and Analysis of Mice Comprising PKDL Gene Disruptions

To investigate the role of PKDL genes, disruptions in PKDL genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in PKDL genes

were created. More particularly, as shown in Figure 11, a PKDL-specific targeting construct having the ability to disrupt a PKDL gene comprising SEQ ID NO:8 was created using in the targeting arms of the construct the oligonucleotide sequences identified herein as SEQ ID NO:10 and SEQ ID NO:11.

The targeting construct was introduced into ES cells derived from the 129/OlaHsd mouse substrain. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

The transgenic mice comprising disruptions in PKDL genes were analyzed for phenotypic changes and expression patterns.

Wild-type control mice, as well as heterozygous and homozygous mutant mice, were evaluated by the following examinations or tests:

- Physical examination.
- Necropsy including body length, body weight, and organ weight measurements.
- Histological examination of tissues and organs.
- Bone marrow section evaluation.
- Complete blood counts and differentials.
- Clinical chemistry panels.
- Fertility evaluation.
- Behavioral tests.
- Aging studies.
- Densitometry.

Homozygous mutant males of approximately 300 days of age exhibited increased body fat percentage, as determined by densitometry, relative to age-matched wild-type control mice. Specifically, homozygous males (n = 2) had on average 23.6% body fat, compared with 15.2% body fat for wild-type control mice (n = 2).

Densitometry measurements were made using a PIXImus™ (GE Lunar; Madison, WI) densitometer, which utilizes Dual Energy X-ray absorptiometry (DEXA or DXA) technology, wherein an x-ray source exposes the entire mouse to a beam of both high and low energy x-rays. The ratio of attenuation of the high and low energies allows the separation of bone from soft tissue, and, from within the tissue samples, lean and fat.

Each mouse is processed on the densitometer following euthanasia. The mouse is placed on a specimen tray, which is coated with an adhesive to prevent accidental repositioning prior to image acquisition. The subject's head is oriented to the left, the body is aligned so the spinal column is straight, and the tail is wrapped around the subject toward the head. The legs are always positioned

slightly away from the body. The specimen tray with the positioned mouse is then placed into the imaging area. The head of the mouse is excluded from the analysis region.

LacZ Reporter Gene Expression. In general, tissues from 7-12 week old heterozygous mutant mice were analyzed for lacZ expression. Organs from heterozygous mutant mice were frozen, sectioned (10 μm), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase, followed by a Nuclear Fast Red counterstaining.

In addition, for brain, wholemount staining was performed. The dissected brain was cut longitudinally, fixed and stained using X-Gal as the substrate for beta-galactosidase. The reaction was stopped by washing the brain in PBS and then fixed in PBS-buffered formaldehyde.

Wild-type control tissues were also stained for lacZ expression to reveal any background or signals due to endogenous beta-galactosidase activity. The following tissues can show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

LacZ (beta-galactosidase) expression was detectable in testis. LacZ expression was not detected in: brain, spinal cord, sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gallbladder, pancreas, kidney, urinary bladder, trachea, larynx, esophagus, thyroid gland, pituitary gland, adrenal glands, salivary glands, tongue, skeletal muscle, skin, and female reproductive systems.

Male Reproductive Systems

Testis

Many spermatogenic cells of the seminiferous tubules showed faint X-Gal staining.

Example 4: Generation and Analysis of Mice Comprising Glutamate Receptor Channel Gamma-2 Gene Disruptions

To investigate the role of glutamate receptor channel gamma-2 genes, disruptions in glutamate receptor channel gamma-2 genes were produced by homologous recombination. Specifically, transgenic mice comprising disruptions in glutamate receptor channel gamma-2 genes were created. More particularly, as shown in Figure 15, a glutamate receptor channel gamma-2-specific targeting construct having the ability to disrupt a glutamate receptor channel gamma-2 gene specifically comprising SEQ ID NO:12 was created using in the targeting arms of the construct the oligonucleotide sequences identified herein as SEQ ID NO:14 or SEQ ID NO:15.

The targeting construct was introduced into ES cells derived from the 129/OlaHsd mouse substrain. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

The transgenic mice comprising disruptions in glutamate receptor channel gamma-2 genes were analyzed for phenotypic changes and expression patterns.

RT-PCR Expression. Total RNA was isolated from the organs or tissues from adult C57BL/6 wild-type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers.

The highest levels of RNA transcripts were detectable in brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord and eye. Lower levels of RNA transcripts were detectable in Harderian glands, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovaries, uterus and white fat.

LacZ Reporter Gene Expression. In general, tissues from 7-12 week old heterozygous mutant mice were analyzed for lacZ expression. Organs from heterozygous mutant mice were frozen, sectioned (10 μm), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase, followed by a Nuclear Fast Red counterstaining.

In addition, for brain, wholemount staining was performed. The dissected brain was cut longitudinally, fixed and stained using X-Gal as the substrate for beta-galactosidase. The reaction was stopped by washing the brain in PBS and then fixed in PBS-buffered formaldehyde.

Wild-type control tissues were also stained for lacZ expression to reveal any background or signals due to endogenous beta-galactosidase activity. The following tissues can show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

LacZ (beta-galactosidase) expression was detectable in brain, spinal cord, sciatic nerve, eye, Harderian glands, aorta, lung, gall bladder, kidney, urinary bladder, trachea, parathyroid gland, pituitary gland, adrenal glands, salivary glands, tongue, male and female reproductive systems. The most striking expression was seen in nervous tissue and smooth muscle cells. LacZ expression was not detected in: heart thymus, spleen, lymph nodes, bone marrow, liver, pancreas, larynx, esophagus, thyroid gland, skeletal muscle and skin.

Brain

In wholemount staining, all regions of the brain including the olfactory bulb showed strong lacZ expression. On frozen sections, strong lacZ expression was detectable throughout forebrain, midbrain and brainstem with practically all cells showing lacZ staining. In the cerebellum, strong lacZ expression was detectable in the granular layer, white matter and peduncles. Basket cells of the

Purkinje cell layer showed X-Gal signals. Many nuclei in the molecular stain displayed X-Gal staining.

Spinal cord

Practically all cells expressed lacZ. Very strong signals were detectable in motor neurons.

Sciatic Nerve

LacZ expression was detectable in the perineurium.

Eyes

LacZ expression was detectable in the retina, with striking signals in the inner nuclear layer, ganglion layer and pigment layer. Further lacZ expression was detectable in the uvea, sclera and optic nerve.

Harderian Glands

Faint lacZ expression was detectable in blood vessels.

Aorta

Faint lacZ expression was detectable in the wall of the aorta.

Lung

LacZ expression was detectable in smooth muscle cells of blood vessels and bronchioles

Gallbladder

Scattered lacZ expression was detectable in the wall of the gallbladder.

Kidney

Strong lacZ expression was detectable in tubule cells of the cortex and blood vessel walls. Faint X-Gal signals were detectable in glomeruli. In medulla scattered X-Gal staining was observed.

Urinary Bladder

LacZ expression was detectable in smooth muscle cells of the muscularis, in epithelial cells of the mucosa and in blood vessel walls.

Trachea

Faint lacZ expression was detectable in myocytes.

Parathyroid Gland

Practically all cells expressed lacZ.

Pituitary Gland

LacZ expression was detectable in pars distalis, pars intermedia and pars nervosa.

Adrenal Glands

X-Gal staining was detectable in capsule and all cells of the medulla.

Salivary Glands

Ganglia in the salivary glands displayed strong lacZ expression.

Tongue

LacZ expression was detectable in blood vessels.

Male Reproductive Systems

Testis

Many spermatogenic cells in the seminiferous tubules expressed lacZ.

Penis

Weak lacZ expression was detectable in myocytes.

Seminal Vesicles

Myocytes in the capsule expressed lacZ strongly.

Coagulating Glands

Myocytes in the capsule expressed lacZ strongly.

Prostate and Ampullary Glands

Myocytes in the capsule expressed lacZ.

Female Reproductive Systems

Ovary

Few follicles showed faint X-Gal staining.

Oviduct/Uterus

Many myocytes of the myometrium displayed X-Gal staining. Very faint signals were detectable in blood vessels.

Vagina/Cervix

LacZ expression was detectable in blood vessels, ganglia and myocytes.

As is apparent to one of skill in the art, various modifications of the above embodiments can be made without departing from the spirit and scope of this invention. These modifications and variations are within the scope of this invention.

CLAIMS

We claim:

1. A targeting construct comprising:

   (a) a first polynucleotide sequence homologous to at least a first portion of a target gene;

   (b) a second polynucleotide sequence homologous to at least a second portion of the target gene; and

   (c) a selectable marker,

   wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.

2. The targeting construct of claim 1, wherein the targeting construct further comprises a screening marker.

3. A method of producing a targeting construct, the method comprising:

   (a) providing a first polynucleotide sequence homologous to at least a first portion of a target gene;

   (b) providing a second polynucleotide sequence homologous to the target gene;

   (c) providing a selectable marker; and

   (d) inserting the first sequence, second sequence, and selectable marker into a vector, to produce the targeting construct,

   wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.

4. A method of producing a targeting construct, the method comprising:

   (a) providing a polynucleotide comprising a first sequence homologous to a first region of a target gene and a second sequence homologous to a second region of the target gene;

   (b) inserting a positive selection marker in between the first and second sequences to form the targeting construct,

   wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.

5. A cell comprising a disruption in a target gene, wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.

6. The cell of claim 5, wherein the cell is a murine cell.

7. The cell of claim 6, wherein the murine cell is an embryonic stem cell.

8. A transgenic mouse comprising a disruption in a target gene, wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.

9. A cell derived from the transgenic mouse of claim 8.

10. A method of producing a transgenic mouse comprising a disruption in a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, the method comprising:

(a) introducing the targeting construct of claim 1 into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse.

11. A method of identifying an agent that modulates the expression of a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, the method comprising:

(a) providing a non-human transgenic animal comprising a disruption in the target gene;

(b) administering an agent to the non-human transgenic animal; and

(c) determining whether the expression of the target gene in the non-human transgenic animal is modulated.

12. A method of identifying an agent that modulates the function of a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, the method comprising:

(a) providing a non-human transgenic animal comprising a disruption in the target gene;

(b) administering an agent to the non-human transgenic animal; and

(c) determining whether the function of the target gene in the non-human transgenic animal is modulated.

13. A method of identifying an agent that modulates the expression of a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, the method comprising:

(a) providing a cell comprising a disruption in the target gene;

(b) contacting the cell with an agent; and

(c) determining whether expression of the target gene is modulated.

14. A method of identifying an agent that modulates the function of a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, the method comprising:

(a) providing a cell comprising a disruption in the target gene;

(b) contacting the cell with an agent; and

(c) determining whether the function of the target gene is modulated.

15. An agent identified by the method of any one of claims 11, 12, 13 or 14.

16. A transgenic mouse comprising a disruption in a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, wherein there is no significant expression of the target gene in the transgenic mouse.
17. A transgenic mouse comprising a disruption in a PKDL gene, wherein the transgenic mouse exhibits increased body fat percentage, relative to a wild-type control mouse.
18. An agonist or antagonist of a target gene product, wherein the target gene is selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene.
19. Phenotypic data associated with a transgenic mouse comprising a disruption in a target gene selected from the group consisting of a Shab potassium channel-like gene, a CCHB3 gene, a PKDL gene, and a glutamate receptor channel gamma-2 gene, wherein the phenotypic data is in an electronic database.

AATTCGGCACGCAGGGGTTCTTCTCATTTCCCCCTACCTCTCTGTCACAACCCTGTCTTTTCCTGGTC
AGTGAGACATGGGCTCCCCATGCCCGGCAGAGAACCAGCTGGGAACACACTTCCTAAAACAAGCGACA
GGTGAGCCCAGTGATCCCTGGCCTGACAGTCTGAGGCAATGCCATGTCTTCCAGAGACAGGGACTTGC
ATCCTGGACACCATCACTTTGGCTCCTGCAGCCCCTTGAGCCAGCTCTGGCCGGGCCCCGAGCCTAAG
TCAGTCAAGGGCCTTTACTACAGCAGGGCCCGGAAGGTGGGCAACCAGGACGCCTCTCCGGAGCCAAC
TTGAAGGAGATCCTAGTGAATGTGGGTGGCCAGCGGTACCTGCTGCCCTGGAGCACCCTGGATGCCTT
CCCGCTGAGCCGCCTGAGCAGGCTCCGGCTGTGCGCCAGCCATGAGGAGATCACGCAGCTCTGGAATG
ACTACGATGAGGACAGCCAGGAGTTCTTCTTCGACAGGAACCCCAGCGCCTTCGGG (SEQ ID
NO:1)

FIGURE 1

underlined = deleted in targeting construct

BOLD = sequence flanking Neo insert in targeting construct

AATTCGGCACGCAGGGGTTCTTCTCATTTCCCCCTACCTCTCTGTCACAACCCTGTCTTT
TCCTGGTCAGTGAGACATGGGCTCCCCATGCCCGGCAGAGAACCAGCTGGGAACACACTT
CCTAAAACAAGCG**ACAGGTGAGCCCAGTGATCCCTGGCCTGACAGTCTGAGGCAATGCCA**
TGTCTTCCAGAGACAGGGACTTGCATCCTGGACACCATCACTTTGGCTCCTGCAGCCCCT
TGAGCCAGCTCTGGCCGGGCCCCGAGCCTAAGTCAGTCAAGGGCCTTTACTACAGCAGGG
CCCGGAAGGTGGGCAACCAGGACGCCTCTCCGGAGCCAACTTGAAGGAGATCCTAGTGAA
TGTGGGTGGCCAGCGGTACCTGCTGCCCTGGAGCACCCTGGATGCCTTCCCGCTGAGCCG
CCTGAGCAGGCTCCGGCTGTGCGCCAG**CCATGAGGAGATCACGCAGCTCTGGAATGACTA**
CGATGAGGACAGCCAGGAGTTCTTCTTCGACAGGAACCCCAGCGCCTTCGGG

FIGURE 2

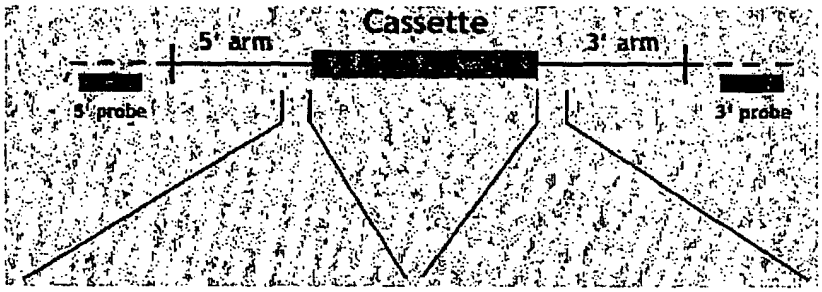
Gene Sequence Structure*

287 bp Sequence Deleted 447 bp

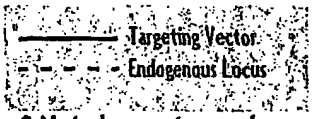
Size of EST: 532 bp

Targeting Vector* (genomic sequence)

LacZ-Neo Cassette



Arm Length:
5': 0.9 kb
3': 4 kb



* Not drawn to scale

5'>GGAGACTCACCCTCCCTTTCT
TCTCTTCTTCTTCCTCTGTTGGCA
GACAGGTGAGCCCAGTGATCCCTG
GCCTGACAGTCTGAGGCAATGCCC
ATGTCTTCCAGAGACAGGGACTTG
CATCCTGGACACCATCACTTTGGC
TCCTGCAGCCCCTTGAGCCAGCTC
TGGCCGGGCCCCGAGCCTAAGTCA
GTCAAGGGCCT<3'
(SEQ ID NO:2)

5'>CCATGAGGAGATCACGCAGCT
CTGCGATGACTACGATGAGGACAG
CCAGGAGTTCTTCTTCGACAGGAA
CCCCAGCGCCTTCGGGGTGATCGT
GAGCTTCCTGGCCGCGGGAAAGCT
GGTGCTTCTGCGAGAGATGTGCGC
CCTGTCCTTCCGGGAGGAGCTGAG
CTACTGGGGCATCGAGGAAACCAA
CCTGGAGCGCT<3'
(SEQ ID NO:3)

FIGURE 3

GGGCGCGGCGCCGTCCTCGCCCGCGCTCGCTCCCCCCACCCACCCCGGACTCCCCCATGTATGACGACTC
CTACGTGCCCGGGTTTGAGGACTCGGAGGCGGGTTCAGCCGACTCCTACACCAGCCGCCCCTCTCTGGAC
TCAGACGTCTCCCTGGAGGAGGACCGGGAGAGTGCCCGGCGAGAAGTGGAGAGTCAGGCTCAGCAGCAGC
TGGAAAGAGCCAAGCACAAACCTGTGGCATTTGCTGTGAGGACCAATGTCAGCTACTGTGGGGTTCTGGA
TGAGGAGTGCCCGGTCCAGGCCTCTGGAGTCAACTTTGAGGCCAAAGATTTTCTGCACATTAAAGAGAAG
TACAGCAATGACTGGTGGATCGGGAGGCTAGTGAAAGAAGGCGGAGCAATCGCCTTCATCCCCAGCCCCC
AACGCCTGGAGAGCATCCGCCTCAAACAGGAACAGAAGGCCAGGAGATCTGGGAACCCTTCCAGCCTGGG
TGACATTGGCTTCCGACGCTCTCCTCCTCCGTCTCTAGCCAAGCAGAAGCAAAAGCAGGCGGAACATGTT
CCCCCGTATGATGTGGTGCCCTCCATGCGGCCTGTGGTGCTGGTGGGACCCTCTCTGAAAGGCTATGAGG
TCACGGACATGATGCAGAAGGCGCTCTTCGACTTCCTTAAACACAGGTTTGATGGCAGGATCTCCATCAC
ACGCGTCACGGCCGACCTCTCGCTGGCCAAGCGCTCTGTGCTCAACAATCCTGGCAAGAGGACCATCATT
GAGCGCTCCTCCGCCCGCTCCAGCATTGCTGAAGTGCAGAGTGAGATTGAGCGCATATTCGAGCTGGCCA
AATCCCTGCAGCTAGTGGTGTTGGATGCTGACACCATCAACCACCCAGCACAACTTGCCAAGACCTCACT
GGCTCCCATCATCGTCTTTGTCAAAGTGTCCTCGCCAAAGGTACTGCAGCGACTGATCCGATCCAGGGGA
AAGTCACAGATGAAGCACCTCACTGTACAGATGATGGCGTACGATAAGCTGGTTCAGTGCCCACCCGAGT
CATTCGACGCGATTCTGGATGAGAACCAGCTAGAAGATGCCTGTGAGCACCTGGCTGAATACTTAGAGGT
TTACTGGCGAGCGACCCACCACCCAGCACCGGGCCCCGGACTGCTGGGTCCGCCTAGTGCCATCCCTGGA
CTTCAGAACCAGCAGCAGCTGGGGGAGCGAGTGGAGGAGCACTCACCCCTGGAGAGGGACAGCCTGATGC
CCTCAGATGAGGCCAGCGAGAGCTCCCGCCAGGCCTGGACCGGATCTTCACAGCGCAGCTCTCGCCATCT
GGAGGAGGACTATGCAGATGCCTACCAGGACCTGTACCAGCCTCACCGCCAACACACCTCGGGGCTGCCC
AGTGCTAACGGGCACGACCCCCAAGACCGGCTCCTAGCCCAGGACTCGGAGCATGACCACAATGACCGGA
ACTGGCAGCGTAACCGGCCTTGGCCCAAGGACAGCTACTGACCACCTCCTGCCCCACCCTGGCAGGCGCA
GGCACAGCGGCTGGGGTGTCCACCTCAGGCAGGTTGGAGTTAGATTGGCATTAGGCTGCCGTTAGTTCAG
CTCACACAACCCTCTGCCCAGCCCCAGGTCCAGGGCTGACTGTGGTCCCAAGGTTCTGGGAGAAGCAAGG
GGCCCCTCACCTCCTGGGCACAGTGACCCCGTAGGTTCTCATCCGGGTACTAGCCGTGTTCTGCATCCTT
GGCACCTCCCCCTGCATAAGCTGCCGCCCCCGTGGGCAACAATCTCAGGCCAGGATCACTTAGCAGGGGC
CTTCCAGCCAGAATGGATGCCCCTCTAAAGAGCAAGAGGGTGTGAGTGTGGGCAACATAGCCTGAGGAAG
AAGAAACTCGGTTCCTAAGCAGGTGTAGATCCTAAGCAAAGGGACTCCATTCACGCCACTGCCACACATC
AGAAATGAAGCAATCAGAGCTCAACATGGCGGCACTTCTGTCCCATCAGCTGGGGTGGGCACTTACACCT
AAGACAGGAGCAGTGCGGGTGAGGCAGGACAGACAGACTCACAGCTGTAGCTCTGCTAGAAAACGGGGGA
CTCAACCAAACCGGGAGGCTTAGCATCTGGTGAGACTGGGGAACTGGGGCATATTCAAGCCAAGAGCCAG
CCTGGACTGGGGGGGAGGGTGGGACAGCTTCCGGCCCCCCTTGCTCTTCTCATTCTTTGCCCTTGCATCT
GTCATTTCTGTCCTTTCCCTCCATGGCTCCTGCAAGATAGGGGCTTCCTGACTCATAGCAGCCACTTCAG
TTAGGGTTAGATGAGAGGAACAGGACACAGTGAACAGCCCCCGAGGCTGTCCACCTGGCTACCCTTGCCT
TATGGCTCTAGCGTGTGACCTACAGAGCATGCTCCATTAAGAACCCGCCCCACCTCATTGTCATCTCCAA
TAAAACACCACGCACAGTC (SEQ ID NO:4)

FIGURE 4

MYDDSYVPGFEDSEAGSADSYTSRPSLDSDVSLEEDRESARREVESQAQQQLERAKHKPVAFAVRTNV
SYCGVLDEECPVQASGVNFEAKDFLHIKEKYSNDWWIGRLVKEGGAIAFIPSPQRLESIRLKQEQKAR
RSGNPSSLGDIGFRRSPPPSLAKQKQKQAEHVPPYDVVPSMRPVVLVGPSLKGYEVTDMMQKALFDFL
KHRFDGRISITRVTADLSLAKRSVLNNPGKRTIIERSSARSSIAEVQSEIERIFELAKSLQLVVLDAD
TINHPAQLAKTSLAPIIVFVKVSSPKVLQRLIRSRGKSQMKHLTVQMMAYDKLVQCPPESFDAILDEN
QLEDACEHLAEYLEVYWRATHHPAPGPGLLGPPSAIPGLQNQQQLGERVEEHSPLERDSLMPSDEASE
SSRQAWTGSSQRSSRHLEEDYADAYQDLYQPHRQHTSGLPSANGHDPQDRLLAQDSEHDHNDRNWQRN
RPWPKDSY (SEQ ID NO:5)

FIGURE 5

underlined = deleted in targeting construct

BOLD = sequence flanking Neo insert in targeting construct

GGGCGCGGCGCCGTCCTCGCCCGCGCTCGCTCCCCCCACCCACCCCGGACTCCCCCATGT
ATGACGACTCCTACGTGCCCGGGTTTGAGGACTCGGAGGCGGGTTCAGCCGACTCCTACA
CCAGCCGCCCCTCTCTGGACTCAGACGTCTCCCTGGAGGAGGACCGGGAGAGTGCCCGGC
GAGAAGTGGAGAGTCAGGCTCAGCAGCAGCTGGAAAGAGCCAAG**CACAAACCTGTGGCAT**
TTGCTGTGAGGACCAATGTCAGCTACTGTGGGGTTCTGGATGAGGAGTGCCCGGTCCAGG
CCTCTGGAGTCAACTTTGAGGCCAAAGATTTTCTGCACATTAAAGAGAAGTACAGCAA**TG**
ACTGGTGGATCGGGAGGCTAGTGAAAGAAGGCGGAGCAATCGCCTTCATCCCCAGCCCCC
AACGCCTGGAGAGCATCCGCCTCAAACAGGAACAGAAGGCCAGGAGATCTGGGAACCCTT
CCAGCCTGGGTGACATTGGCTTCCGACGCTCTCCTCCTCCGTCTCTAGCCAAGCAGAAGC
AAAAGCAGGCGGAACATGTTCCCCCGTATGATGTGGTGCCCTCCATGCGGCCTGTGGTGC
TGGTGGGACCCTCTCTGAAAGGCTATGAGGTCACGGACATGATGCAGAAGGCGCTCTTCG
ACTTCCTTAAACACAGGTTTGATGGCAGGATCTCCATCACACGCGTCACGGCCGACCTCT
CGCTGGCCAAGCGCTCTGTGCTCAACAATCCTGGCAAGAGGACCATCATTGAGCGCTCCT
CCGCCCGCTCCAGCATTGCTGAAGTGCAGAGTGAGATTGAGCGCATATTCGAGCTGGCCA
AATCCCTGCAGCTAGTGGTGTTGGATGCTGACACCATCAACCACCCAGCACAACTTGCCA
AGACCTCACTGGCTCCCATCATCGTCTTTGTCAAAGTGTCCTCGCCAAAGGTACTGCAGC
GACTGATCCGATCCAGGGGAAAGTCACAGATGAAGCACCTCACTGTACAGATGATGGCGT
ACGATAAGCTGGTTCAGTGCCCACCCGAGTCATTCGACGCGATTCTGGATGAGAACCAGC
TAGAAGATGCCTGTGAGCACCTGGCTGAATACTTAGAGGTTTACTGGCGAGCGACCCACC
ACCCAGCACCGGGCCCCGGACTGCTGGGTCCGCCTAGTGCCATCCCTGGACTTCAGAACC
AGCAGCAGCTGGGGGAGCGAGTGGAGGAGCACTCACCCCTGGAGAGGGACAGCCTGATGC
CCTCAGATGAGGCCAGCGAGAGCTCCCGCCAGGCCTGGACCGGATCTTCACAGCGCAGCT
CTCGCCATCTGGAGGAGGACTATGCAGATGCCTACCAGGACCTGTACCAGCCTCACCGCC
AACACACCTCGGGGCTGCCCAGTGCTAACGGGCACGACCCCCAAGACCGGCTCCTAGCCC
AGGACTCGGAGCATGACCACAATGACCGGAACTGGCAGCGTAACCGGCCTTGGCCCAAGG
ACAGCTACTGACCACCTCCTGCCCCACCCTGGCAGGCGCAGGCACAGCGGCTGGGGTGTC
CACCTCAGGCAGGTTGGAGTTAGATTGGCATTAGGCTGCCGTTAGTTCAGCTCACACAAC
CCTCTGCCCAGCCCCAGGTCCAGGGCTGACTGTGGTCCCAAGGTTCTGGGAGAAGCAAGG
GGCCCCTCACCTCCTGGGCACAGTGACCCCGTAGGTTCTCATCCGGGTACTAGCCGTGTT
CTGCATCCTTGGCACCTCCCCCTGCATAAGCTGCCGCCCCCGTGGGCAACAATCTCAGGC
CAGGATCACTTAGCAGGGGCCTTCCAGCCAGAATGGATGCCCCTCTAAAGAGCAAGAGGG
TGTGAGTGTGGGCAACATAGCCTGAGGAAGAAGAAACTCGGTTCCTAAGCAGGTGTAGAT
CCTAAGCAAAGGGACTCCATTCACGCCACTGCCACACATCAGAAATGAAGCAATCAGAGC
TCAACATGGCGGCACTTCTGTCCCATCAGCTGGGGTGGGCACTTACACCTAAGACAGGAG
CAGTGCGGGTGAGGCAGGACAGACAGACTCACAGCTGTAGCTCTGCTAGAAAACGGGGGA
CTCAACCAAACCGGGAGGCTTAGCATCTGGTGAGACTGGGGAACTGGGGCATATTCAAGC
CAAGAGCCAGCCTGGACTGGGGGGGAGGGTGGGACAGCTTCCGGCCCCCCTTGCTCTTCT
CATTCTTTGCCCTTGCATCTGTCATTTCTGTCCTTTCCCTCCATGGCTCCTGCAAGATAG
GGGCTTCCTGACTCATAGCAGCCACTTCAGTTAGGGTTAGATGAGAGGAACAGGACACAG
TGAACAGCCCCCGAGGCTGTCCACCTGGCTACCCTTGCCTTATGGCTCTAGCGTGTGACC
TACAGAGCATGCTCCATTAAGAACCCGCCCCACCTCATTGTCATCTCCAATAAAACACCA
CGCACAGTC

FIGURE 6

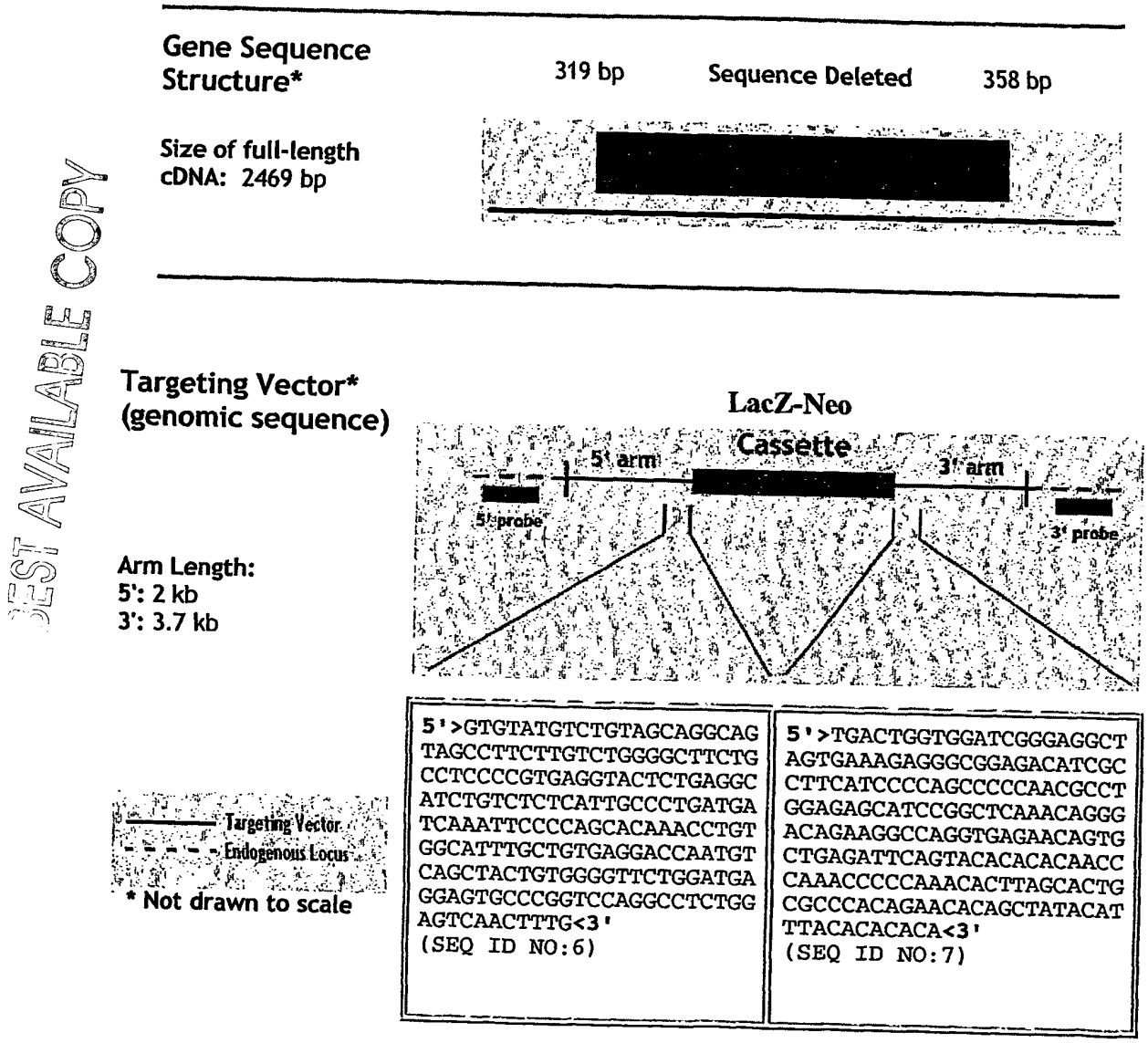


FIGURE 7

CAGCATCGCCTTTTCCTCCCGTTTCTCCTTCCACTCCCAGCTCCACATCCTCCTCCTATTCTCCCCTCTC
CCCTCTTCAAACCCCCACCTTCCAGTTCCCTCACCTCCCCTTTCGGCTGGTCCCCTGGGGCTTGCAGCAA
GAGGGAGAGAGAGCTCCTGACAGGATTGATGGTCCTTCCCCACCCTGTCCTCTCATCCGCTCCCTCCCCA
GCAGGCACAGACATCCCCCTACAAAAGGCAGGAGCCCAGGCTGTGTGGAAACAGCTGCTCTCAGACGCCT
TTCCATTTGCTCTCTGCTGGCTAGGCTGGGCTGTGCCTCTGCTCCCTCTTCCTCTAGCTGAGAGTGGGCA
CCTGGGGTACCGGGCCCCCCCACCTCATTCCCCATGAATGCTGTGGGAAGTCCTGAGGGGCAGGAGCTGC
ACAAGCTGGGGAGTGGAGCCTGGGACAACCCCGCCTACAGTGGTCCCCCTTCCCCACACGGGACGCTGAG
AGTCTGCACCATCTCCAGCACGGGGCCTCTCCAGCCCCAACCCAAGAAGCCTGAAGATGAACCCCAGGAG
ACGGCATACAGGACCCAGGTGTCCAGCTGCTGCCTCCATATCTGTCAAGGCATCAGAGGACTTTGGGGAA
CAACCCTGACTGAGAACACAGCTGAGAACCGGGAACTTTATATCAAGACCACCCTGAGGGAGCTGTTGGT
ATATATTGTGTTCCTGGTGGACATCTGTCTACTGACCTATGGAATGACAAGCTCCAGTGCTTATTACTAC
ACCAAAGTGATGTCTGAGCTCTTCTTACATACTCCATCAGACACTGGAGTCTCCTTTCAGGCCATCAGCA
GCATGGCGGACTTCTGGGATTTTGCCCAGGGCCCACTACTGGACAGTTTGTATTGGACCAAATGGTACAA
CAACCAGAGCCTGGGCCATGGCTCCCACTCCTTCATCTACTATGAGAACATGCTGCTGGGGGTTCCGAGG
CTGCGGCAGCTAAAGGTCCGCAATGACTCCTGTGTGGTGCATGAAGACTTCCGGGAGGACATTCTGAGCT
GCTATGATGTCTACTCTCCAGACAAAGAAGAACAACTCCCCTTTGGGCCCTTCAATGGCACAGCGTGGAC
ATACCACTCGCAGGATGAGTTGGGGGGCTTCTCCCACTGGGGCAGGCTCACAAGCTACAGCGGAGGTGGC
TACTACCTGGACCTTCCAGGATCCCGACAGGGTAGTGCAGAGGCTCTCCGGGCCCTTCAGGAGGGGCTGT
GGCTGGACAGGGGCACTCGAGTGGTGTTCATCGACTTCTCAGTCTACAATGCCAATATCAATCTTTTCTG
TGTCCTGAGGCTGGTGGTGGAGTTTCCAGCTACAGGAGGTGCCATCCCATCCTGGCAAATCCGCACAGTC
AAGCTGATCCGCTATGTCAGCAACTGGGACTTCTTTATCGTTGGCTGTGAGGTCATCTTCTGCGTCTTCA
TCTTCTACTATGTGGTGGAAGAGATCCTGGAGCTCCACATTCACCGGCTTCGCTACCTCAGCAGCATCTG
GAACATACTGGACCTGGTGATCATCTTGCTCTCCATTGTGGCTGTGGGCTTCCACATATTCCGAACCCTC
GAGGTGAATCGGCTCATGGGGAAGCTCCTGCAGCAGCCAAACACGTATGCAGACTTTGAGTTCCTCGCCT
TCTGGCAGACACAGTACAACAACATGAATGCTGTCAACCTCTTCTTCGCCTGGATCAAGATATTCAAGTA
CATCAGCTTCAACAAAACCATGACCCAGCTCTCCTCCACGCTGGCCCGCTGTGCCAAGGACATCCTGGGC
TTCGCCGTCATGTTCTTCATTGTTTTCTTCGCCTATGCCCAACTCGGCTACCTGCTTTTCGGGACCCAAG
TGGAAAACTTTAGCACTTTCATCAAGTGCATTTTCACTCAGTTCCGGATAATCCTCGGGGACTTTGACTACAATG
CTATCGACAATGCCAACCGCATCCTGGGCCCTGCCTACTTTGTCACCTATGTCTTCTTCGTCTTCTTCGTGCTCC
TGAACATGTTCCTGGCCATCATCAATGACACATATTCAGAGGTCAAGGAGGAGCTGGCTGGACAGAAGGATGAGC
TGCAACTTTCTGACCTCCTGAAACAGGGCTACAACAAGACCCTACTAAGACTGCGTCTGAGGAAGGAGAGGGTTT
CGGATGTGCAGAAGGTCCTGCAGGGTGGGGAGCAGGAGATCCAGTTTGAGGATTTCACCAACACCTTAAGGGAAC
TGGGACACGCAGAGCATGAAATCACTGAGCTCACGGCCACCTTCACCAAGTTTGACAGAGATGGGAATCGTATTC
TGGATGAGAAGGAACAGGAAAAAATGCGACAGGACCTGGAGGAAGAGAGGGTGGCCCTCAACACTGAGATTGAGA
AACTAGGCCGATCTATTGTGAGCAGCCCACAAGGCAAATCGGGTCCAGAGGCTGCCAGAGCAGGAGGCTGGGTTT
CAGGAGAAGAATTCTACATGCTCACAAGGAGAGTTCTGCAGCTGGAGACTGTCCTGGAAGGAGTAGTGTCCCAGA
TTGATGCTGTAGGCTCAAAGCTGAAAATGCTGGAGAGGAAGGGGTGGCTGGCTCCCTCCCCAGGCGTGAAGGAAC
AAGCTATTTGGAAGCACCCGCAGCCAGCCCCAGCTGTGACCCCAGACCCCTGGGGAGTCCAGGGTGGGCAGGAGA
GTGAGGTTCCCTATAAAAGAGAAGAGGAAGCCTTAGAGGAGAGGAGACTCTCCCGTGGTGAGATTCCAACGTTGC
AGAGGAGTTAAGTGTGAGGCACTCCCGGAGCAAAGTCTATGAAGGATCTTCTGCAAGAGGCTGCCTCCTGGTCCA
CTGAACCTGGAAACTGAGTGGGCTTTAACCAGGAGATAAAAATGGAGCCTGAAGGGAATCAGGCAAGGAAATGAA
CTCAGGATTCAGAGATCTTTGAATTAATATGTGGTGGGTTCTGACATTATTCTTCCATAAGACCATGTGGGTTTC
CATGGTGGCTATCAATAAAACTCCTTAGT (SEQ ID NO:8)

FIGURE 8

MNAVGSPEGQELHKLGSGAWDNPAYSGPPSPHGTLRVCTISSTGPLQPQPKKPEDEPQETAYRTQVSSCCLHICQ
GIRGLWGTTLTENTAENRELYIKTTLRELLVYIVFLVDICLLTYGMTSSSAYYYTKVMSELFLHTPSDTGVSFQA
ISSMADFWDFAQGPLLDSLYWTKWYNNQSLGHGSHSFIYYENMLLGVPRLRQLKVRNDSCVVHEDFREDILSCYD
VYSPDKEEQLPFGPFNGTAWTYHSQDELGGFSHWGRLTSYSGGGYYLDLPGSRQGSAEALRALQEGLWLDRGTRV
VFIDFSVYNANINLFCVLRLVVEFPATGGAIPSWQIRTVKLIRYVSNWDFFIVGCEVIFCVFIFYYVVEEILELH
IHRLRYLSSIWNILDLVIILLSIVAVGFHIFRTLEVNRLMGKLLQQPNTYADFEFLAFWQTQYNNMNAVNLFFAW
IKIFKYISFNKTMTQLSSTLARCAKDILGFAVMFFIVFFAYAQLGYLLFGTQVENFSTFIKCIFTQFRIILGDFD
YNAIDNANRILGPAYFVTYVFFVFFVLLNMFLAIINDTYSEVKEELAGQKDELQLSDLLKQGYNKTLLRLRLRKE
RVSDVQKVLQGGEQEIQFEDFTNTLRELGHAEHEITELTATFTKFDRDGNRILDEKEQEKMRQDLEEERVALNTE
IEKLGRSIVSSPQGKSGPEAARAGGWVSGEEFYMLTRRVLQLETVLEGVVSQIDAVGSKLKMLERKGWLAPSPGV
KEQAIWKHPQPAPAVTPDPWGVQGGQESEVPYKREEEALEERRLSRGEIPTLQRS (SEQ ID NO:9)

FIGURE 9

underlined = deleted in targeting construct

BOLD = sequence flanking Neo insert in targeting construct

CAGCATCGCCTTTTCCTCCCGTTTCTCCTTCCACTCCCAGCTCCACATCCTCCTCCTATT
CTCCCCTCTCCCCTCTTCAAACCCCCACCTTCCAGTTCCCTCACCTCCCCTTTCGGCTGG
TCCCCTGGGGCTTGCAGCAAGAGGGAGAGAGAGCTCCTGACAGGATTGATGGTCCTTCCC
CACCCTGTCCTCTCATCCGCTCCCTCCCCAGCAGGCACAGACATCCCCCTACAAAAGGCA
GGAGCCCAGGCTGTGTGGAAACAGCTGCTCTCAGACGCCTTTCCATTTGCTCTCTGCTGG
CTAGGCTGGGCTGTGCCTCTGCTCCCTCTTCCTCTAGCTGAGAGTGGGCACCTGGGGTAC
CGGGCCCCCCCACCTCATTCCCCATGAATGCTGTGGGAAGTCCTGAGGGGCAGGAGCTGC
ACAAGCTGGGGAGTGGAGCCTGGGACAACCCCGCCTACAGTGGTCCCCCTTCCCCACACG
GGACGCTGAGAGTCTGCACCATCTCCAGCACGGGGCCTCTCCAGCCCCAACCCAAGAAGC
CTGAAGATGAACCCCAGGAGACGGCATACAGGACCCAGGTGTCCAGCTGCTGCCTCCATA
TCTGTCAAGGCATCAGAGGACTTTGGGGAACAACCCTGACTGAGAACACAGCTGAGAACC
GGGAACTTTATATCAAGACCACCCTGAGGGAGCTGTTGGTATATATTGTGTTCCTGGTGG
ACATCTGTCTACTGACCTATGGAATGACAAGCTCCAGTGCTTATTACTACACCAAAGTGA
TGTCTGAGCTCTTCTTACATACTCCATCAGACACTGGAGTCTCCTTTCAGGCCATCAGCA
GCATGGCGGACTTCTGGGATTTTGCCCAGGGCCCACTACTGGACAGTTTGTATTGGACCA
AATGGTACAACAACCAGAGCCTGGGCCATGGCTCCCACTCCTTCATCTACTATGAGAACA
TGCTGCTGGGGGTTCCGAGGCTGCGGCAGCTAAAGGTCCGCAATGACTCCTGTGTGGTGC
ATGAAGACTTCCGGGAGGACATTCTGAGCTGCTATGATGTCTACTCTCCAGACAAAGAAG
AACAACTCCCCTTTGGGCCCTTCAATGGCACAGCGTGGACATACCACTCGCAGGATGAGT
TGGGGGGCTTCTCCCACTGGGGCAGGCTCACAAGCTACAGCGGAGGTGGCTACTACCTGG
ACCTTCCAGGATCCCGACAGGGTAGTGCAGAGGCTCTCCGGGCCCTTCAGGAGGGGCTGT
GGCTGGACAGGGGCACTCGAGTGGTGTTCATCGACTTCTCAGTCTACAATGCCAATATCA
ATCTTTTCTGTGTCCTGAGGCTGGTGGTGGAGTTTCCAGCTACAGGAGGTGCCATCCCAT
CCTGGCAAATCCGCACAGTCAAGCTGATCCGCTATGTCAGCAACTGGGACTTCTTTATCG
TTGGCTGTGAGGTCATCTTCTGCGTCTTCATCTTCTACTATGTGGTGGAAGAGATCCTGG
AGCTCCACATTCACCGGCTTCGCTACCTCAGCAGCATCTGGAACATACTGGACCTGGTGA
TCATCTTGCTCTCCATTGTGGCTGTGGGCTTCCACATATTCCGAACCCTCGAGGTGAATC
GGCTCATGGGGAAGCTCCTGCAGCAGCCAAACACGTATGCAGACTTTGAGTTCCTCGCCT
TCTGGCAGACACAGTACAACAACATGAATGCTGTCAACCTCTTCTTCGCCTGGATCAAGA
TATTCAAGTACATCAGCTTCAACAAAACCATGACCCAGCTCTCCTCCACGCTGGCCCGCT
GTGCCAAGGACATCCTGGGCTTCGCCGTCATGTTCTTCATTGTTTTCTTCGCCTATGCCC
AACTCGGCTACCTGCTTTTCGGGACCCAAGTGGAAAACTTTAGCACTTTCATCAAGTGCA
TTTTCACTCAGTTCCGGATAATCCTCGGGGACTTTGACTACAATGCTATCGACAATGCCA
ACCGCATCCTGGGCCCTGCCTACTTTGTCACCTATGTCTTCTTCGTCTTCTTCGTGCTCC
TGAACAT**GTTCCTGGCCATCATCAATGACACATATTCAGAGGTCAAGGAGGAGCTGGCTG**
GACAGAAGGATGAGCTGCAACTTTCTGACCTCCTGAAACAGGGCTACAACAAGACCCTAC
TAAGACTGCGTCTGAGGAAGGAGAGGGTTTCGGATGTGCAGAAGGTCCTGCAGGGTGGGG
AGCAGGAGATCCAGTTTGAGGATTTCACCAACACCTTAAGGGAACTGGGACACGCAGAGC
ATGAAATCACTGAGCTCACGGCCACCTTCACCAAGTTTGACAGAGATGGGAATCGTATTC
TGGATGAGAAGGAACAGGAAAAAATGCGACAGGACCTGGAGGAAGAGAGGGTGGCCCTCA
ACACTGAGATTGAGAAACTAGGCCGATCTATTGTGAGCAGCCCACAAGGCAAATCGGGTC
CAGAGGCTGCCAGAGCAGGAGGCTGGGTTTCAGGAGAAGAATTCTACATGCTCACAAGGA
GAGTTCTGCAGCTGGAGACTGTCCTGGAAGGAGTAGTGTCCCAGATTGATGCTGTAGGCT
CAAAGCTGAAAATGCTGGAGAGGAAGGGGTGGCTGGCTCCCTCCCCAGGCGTGAAGGAAC
AAGCTATTTGGAAGCACCCGCAGCCAGCCCCAGCTGTGACCCCAGACCCCTGGGGAGTCC
AGGGTGGGCAGGAGAGTGAGGTTCCCTATAAAAGAGAAGAGGAAGCCTTAGAGGAGAGGA
GACTCTCCCGTGGTGAGATTCCAACGTTGCAGAGGAGTTAAGTGTGAGGCACTCCCGGAG
CAAAGTCTATGAAGGATCTTCTGCAAGAGGCTGCCTCCTGGTCCACTGAACCTGGAAACT
GAGTGGGCTTTAACCAGGAGATAAAAATGGAGCCTGAAGGGAATCAGGCAAGGAAATGAA
CTCAGGATTCAGAGATCTTTGAATTAATATGTGGTGGGTTCTGACATTATTCTTCCATAA
GACCATGTGGGTTTCCATGGTGGCTATCAATAAAACTCCTTAGT

FIGURE 10

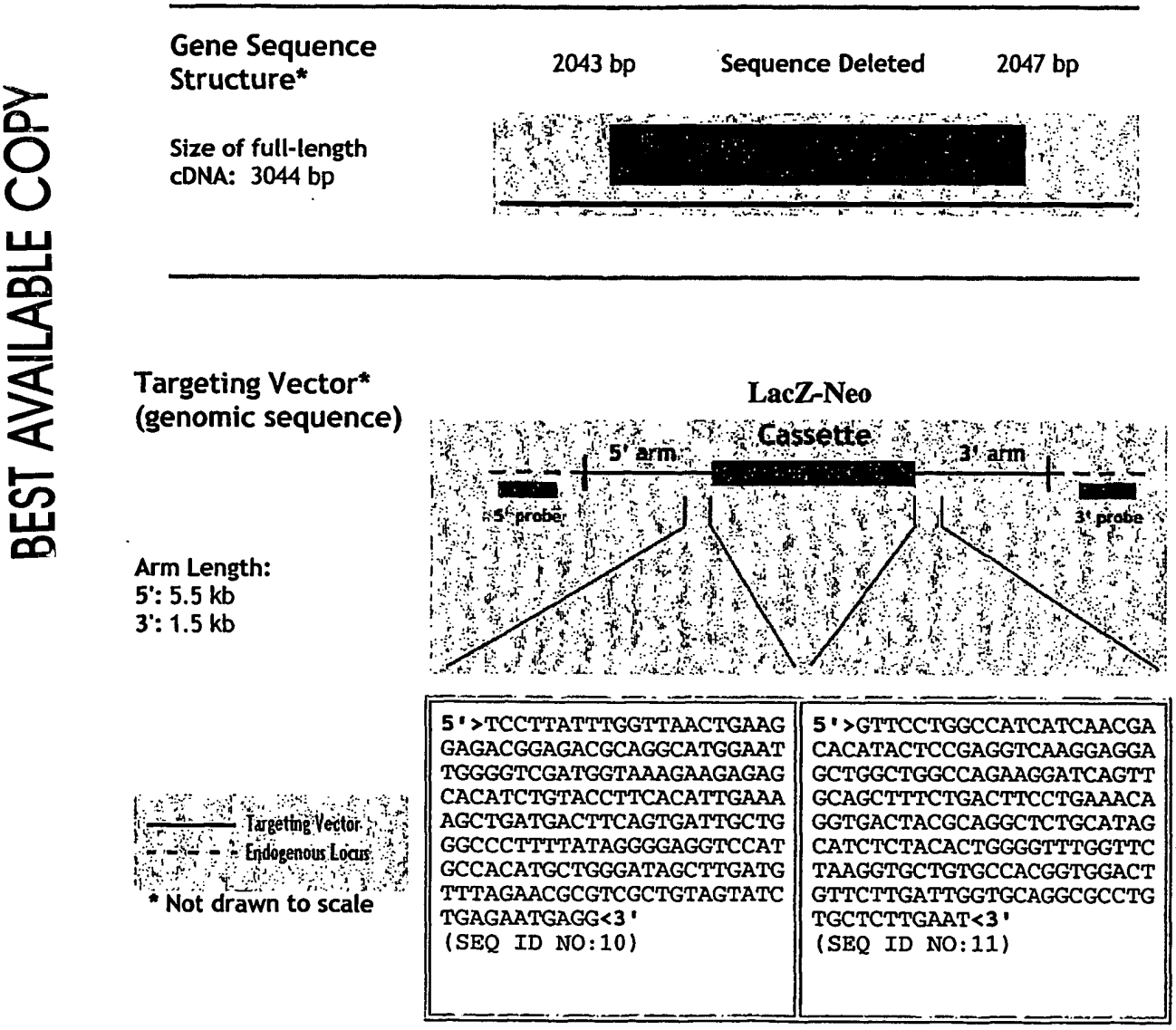


FIGURE 11

CCCCTTCTCTCCTCCATTTTCTCCTTTCCCCTTCCTCCCCTCCTGGCCCGCCCTTTGCTTTTCCACCTG
TCCTTCCTCCCCGCTGCTGGGGAGACACCGTTTTGGATTAGTTGGGGGCCACCGCCGGCGGCGGGGGAG
GGGGCCCAGTGGACTGCCCTCTCCCCCGGCCCCAGCCTCACCAGCACCCAGCATCGGAGCACCTCCTGG
TTGTCGTTCCTCGGGCCTCGAGGGAGCCCAGCCCTCCGTCCCACCAGGATCCGTGGCGAGTGGGGGCCG
CGGCAGCTGCGTCCCCATGAGGAGGGGAGGAAGATGCCGGCTGAGCTGCTGCTGCTGCTGATAGTCGCC
TTCGCCAATCCCAGCTGCCAGGTGCTGTCATCACTGCGCATGGCTGCAATCCTGGACGACCAGACCGTG
TGTGGCCGTGGTGAGCGTCTGGCCCTGGCCCTGGCCCGAGAGCAGATCAATGGGATCATCGAGGTCCCA
GCCAAGGCCAGAGTGGAAGTAGACATCTTTGAGCTGCAGCGGGACAGCCAGTACGAGACCACGGACACC
ATGTGTCAGATCCTGCCCAAGGGGGTTGTATCTGTCTTGGGACCCTCCTCCAGCCCAGCTTCTGCCTCC
ACCGTGAGCCATATCTGTGGGGAGAAGGAGATTCCCCACATCAAGGTGGGTCCTGAGGAGACGCCCCGC
CTTCAGTACCTTCGCTTCGCATCTGTCAGCCTGTACCCCAGTAATGAAGATGTCAGCCTGGCAGTCTCC
CGAATCCTCAAGTCCTTTAACTACCCCTCAGCTAGCCTCATCTGCGCCAAGGCTGAGTGCCTGCTGCGG
CTAGAAGAACTGGTGCGAGGCTTCCTCATCTCCAAGGAGACACTGTCCGTGAGGATGCTTGATGACAGC
CGGGACCCCACGCCGCTACTCAAGGAGATCCGAGATGACAAAGTGTCCACCATCATCATTGATGCCAAT
GCGTCCATCTCCCACCTTGTCCTCCGTAAGGCTTCGGAGCTGGGAATGACCTCAGCGTTTTACAAGTAC
ATCCTCACCACCATGGACTTTCCCATCCTGCATCTGGATGGTATCGTGGAGGACTCCTCCAACATCCTG
GGCTTTTCCATGTTCAACACCTCCCACCCCTTCTACCCAGAGTTTGTGCGCAGCCTCAACATGTCCTGG
AGGGAGAACTGTGAAGCCAGCACCTATCCTGGCCCTGCGCTGTCCGCAGCCCTGATGTTTGACGCTGTG
CACGTGGTGGTAAGCGCTGTCCGAGAACTGAACCGAAGCCAGGAGATTGGCGTCAAGCCACTGGCCTGC
ACTTCGGCCAACATTTGGCCCCATGGGACCAGCCTTATGAACTACCTTCGAATGGTAGAGTATGACGGG
CTGACCGGGCGGGTTGAGTTCAACAGCAAAGGGCAGAGGACCAACTACACACTACGCATCCTGGAGAAG
TCCCGCCAGGGCCACCGTGAGATAGGGGTGTGGTACTCTAACCGGACCCTGGCCATGAATGCCACCACC
CTGGACATCAACCTGTCACAGACTCTAGCCAACAAGACTCTGGTGGTCACAACTATCCTGGAGAACCCG
TATGTTATGCGCCGGCCCAACTTCCAGGCCTTGTCAGGGAATGAGCGCTTCGAGGGCTTCTGCGTGGAC
ATGCTCAGGGAGCTGGCCGAGCTGCTGCGCTTCCGATACCGCCTGCGGTTGGTAGAGGACGGACTCTAC
GGGGCACCTGAGCCCAACGGTTCCTGGACAGGCATGGTTGGAGAACTCATCAACCGGAAGGCAGACCTG
GCTGTGGCAGCCTTCACCATCACCGCCGAGAGGGAGAAGGTCATCGACTTCTCCAAGCCCTTCATGACC
CTGGGGATCAGCATCCTCTACAGGGTGCACATGGGCCGCAAGCCTGGCTACTTCTCCTTCCTGGACCCC
TTCTCCCCTGCCGTGTGGCTCTTCATGCTTCTTGCCTACCTGGCTGTCAGCTGTGTCTTGTTCCTGGCT
GCCAGGCTGAGCCCTTATGAGTGGTACAACCCACACCCGTGTCTCCGGGCGCGTCCCCATATCCTGGAG
AACCAGTACACGCTGGGCAACAGCCTCTGGTTCCCCGTGGGTGGCTTCATGCAGCAGGGCTCGGAGGTC
ATGCCGCGGGCACTGTCCACACGCTGTGTCAGCGGAGTCTGGTGGGCCTTCACCTTGATCATCATCTCC
TCCTACACGGCCAACCTGGCTGCCTTCCTCACGGTGCAGCGCATGGAGGTGCCGGTGGAGTCGGCTGAC
GACCTGGCGGATCAGACCAACATTGAGTACGGCACTATCCACGCTGGCTCCACCATGACCTTCTTCCAG
AACTCGCGGTACCAGACGTACCAGCGGATGTGGAACTACATGCAATCGAAGCAGCCCAGCGTGTTTGTC
AAGAGCACAGAGGAGGGAATCGCCCGCGTCCTCAACTCCCGCTATGCCTTCCTGCTGGAGTCCACCATG
AACGAGTACCACAGGCGCCTCAATTGCAACCTCACCCAGATCGGGGGCCTCCTCGACACCAAGGGCTAC
GGCATCGGCATGCCGCTGGGCTCCCCTTTCCGGGATGAGATCACACTGGCCATCCTGCAGCTCCAGGAG
AACAACAGGCTGGAGATCCTGAAGCGCAAGTGGTGGGAGGGCGGCCGGTGCCCCAAGGAGGAGGACCAC
AGGGCCAAAGGTTTGGGCATGGAGAACATTGGCGGCATTTTTGTCGTGCTGATCTGTGGCCTCATCATT
GCTGTCTTCGTGGCGGTCATGGAGTTCATCTGGTCCACGCGGAGGTCAGCGGAGTCCGAGGAGGTGTCG
GTGTGCCAGGAGATGCTGCAGGAGCTACGCCACGCCGTGTCTTGCCGAAAGACCTCGCGTTCCCGCCGG
CGCCGGCGCCCTGGTGGCCCGAGCCGGGCCCTGCTGTCGCTGCGCGCAGTCCGCGAGATGCGACTCAGC
AACGGCAAGCTCTACTCGGCCGGCGCGGGCGGGGACGCGGGCGCGCACGGGGGTCCGCAGCGCCTCCTG
GACGACCCCGGACCTCCTGGGGGACCCCGGCCCCAGGCTCCCACGCCCTGCACGCACGTGCGCGTCTGC
CAGGAGTGCAGGCGCATCCAGGCGCTGCGAGCTTCGGGGGCCGGGGCGCCCCCACGTGGCCTGGGCACC
CCAGCCGAAGCCACCAGCCCGCCTCGGCCGCGGCCAGGCCCCACCGGACCCCGCGAGCTGACCGAGCAC
GAATGACCGTGGACGGGGCCGGGCGTGCGCCGACTGACTGCAGGGACGCGATGCGCCCCAGACGGACAG
GACGCCGCGATTTTGCCTTCAATTCCTGGTGAAGTCCGAGCCCGGCTCCGGAGCAGGCCTGCGCCTCCT
AGTGGACTTGAGCAAGGGTGTCGCGGACGCCGCATTCTATCCGCACCGTGGCGGAGGAGCGCAGAGACC
GAGGACTCTAAGGGCCGCGGACTGTGGGGGCTGCTCCCGGAGTTGGAAAGCGGTCCGCGGGAGGACCCC
AACCTGGGACTGCCCAGATGCCCCAAGACTTGACACGGCTCTCCACACTTCTGGAGGTGGGGAGGGCTC
TGGACATGTGGGTGTACCCTGTGCCCCTCTCTTCTCTTTCTCTCTTTTTTGGGGGGAGAAACCTCAGAA
ATTTCTATGAGACGCCCCCAGGGAGAGGGTCTCTTGGGCCCCTATACCTCCCCCTGCCACATCCCAGTC
CTGTTGG (SEQ ID NO:12)

FIGURE 12

mpaellllli vafanpscqv lsslrmaail ddqtvcgrge rlalalareq ingiievpak
arvevdifel qrdsqyettd tmcqilpkgv vsvlgpsssp asastvshic gekeiphikv
gpeetprlqy lrfasvslyp snedvslavs rilksfnyps aslicakaec llrleelvrg
flisketlsv rmlddsrdpt pllkeirddk vstiiidana sishlvlrka selgmtsafy
kyilttmdfp ilhldgived ssnilgfsmf ntshpfypef vrslnmswre nceastypgp
alsaalmfda vhvvvsavre lnrsqeigvk plactsaniw phgtslmnyl rmveydgltg
rvefnskgqr tnytlrilek srqghreigv wysnrtlamn attldinlsq tlanktlvvt
tilenpyvmr rpnfqalsgn erfegfcvdm lrelaellrf ryrlrlvedg lygapepngs
wtgmvgelin rkadlavaaf titaerekvi dfskpfmtlg isilyrvhmg rkpgyfsfld
pfspavwlfm llaylavscv lflaarlspy ewynphpclr arphilenqy tlgnslwfpv
ggfmqqgsev mpralstrcv sgvwwaftli iissytanla afltvqrmev pvesaddlad
qtnieygtih agstmtffqn sryqtyqrmw nymqskqpsv fvksteegia rvlnsryafl
lestmneyhr rlncnltqig glldtkgygi gmplgspfrd eitlailqlq ennrleilkr
kwweggrcpk eedhrakglg meniggifvv licgliiavf vavmefiwst rrsaeseevs
vcqemlqelr havscrktsr srrrrrpggp srallslrav remrlsngkl ysagaggdag
ahggpqrlld dpgppggprp qaptpcthvr vcqecrriqa lrasgagapp rglgtpaeat
spprprpgpt gpreltehe (SEQ ID NO:13)

FIGURE 13

underlined = deleted in targeting construct

BOLD = sequence flanking Neo insert in targeting construct

CCCCTTCTCTCCTCCATTTTCTCCTTTCCCCTTCCTCCCCTCCTGGCCCGCCCTTTGCTTTTC
CACCTGTCCTTCCTCCCCGCTGCTGGGGAGACACCGTTTTGGATTAGTTGGGGGCCACCGCCG
GCGGCGGGGGAGGGGGCCCAGTGGACTGCCCTCTCCCCCGGCCCCAGCCTCACCAGCACCCAG
CATCGGAGCACCTCCTGGTTGTCGTTCCTCGGGCCTCGAGGGAGCCCAGCCCTCCGTCCCACC
AGGATCCGTGGCGAGTGGGGGCCGCGGCAGCTGCGTCCCCATGAGGAGGGGAGGAAGATGCCG
GCTGAGCTGCTGCTGCTGCTGATAGTCGCCTTCGCCAATCCCAGCTGCCAGGTGCTGTCATCA
CTGCGCATGGCTGCAATCCTGGACGACCAGACCGTGTGTGGCCGTGGTGAGCGTCTGGCCCTG
GCCCTGGCCCGAGAGCAGATCAATGGGATCATCGAGGTCCCAGCCAAGGCCAGAGTGGAAGTA
GACATCTTTGAGCTGCAGCGGGACAGCCAGTACGAGACCACGGACACCATGTGTCAGATCCTG
CCCAAGGGGGTTGTATCTGTCTTGGGACCCTCCTCCAGCCCAGCTTCTGCCTCCACCGTGAGC
CATATCTGTGGGGAGAAGGAGATTCCCCACATCAAGGTGGGTCCTGAGGAGACGCCCCGCCTT
CAGTACCTTCGCTTCGCATCTGTCAGCCTGTACCCCAGTAATGAAGATGTCAGCCTGGCAGTC
TCCCGAATCCTCAAGTCCTTTAACTACCCCTCAGCTAGCCTCATCTGCGCCAAGGCTGAGTGC
CTGCTGCGGCTAGAAGAACTGGTGCGAGGCTTCCTCATCTCCAAGGAGACACTGTCCGTGAGG
ATGCTTGATGACAGCCGGGACCCCACGCCGCTACTCAAGGAGATCCGAGATGACAAAGTGTCC
ACCATCATCATTGATGCCAATGCGTCCATCTCCCACCTTGTCCTCCGTAAGGCTTCGGAGCTG
GGAATGACCTCAGCGTTTTACAAGTACATCCTCACCACCATGGACTTTCCCATCCTGCATCTG
GATGGTATCGTGGAGGACTCCTCCAACATCCTGGGCTTTTCCATGTTCAACACCTCCCACCCC
TTCTACCCAGAGTTTGTGCGCAGCCTCAACATGTCCTGGAGGGAGAACTGTGAAGCCAGCACC
TATCCTGGCCCTGCGCTGTCCGCAGCCCTGATGTTTGACGCTGTGCACGTGGTGGTAAGCGCT
GTCCGAGAACTGAACCGAAGCCAGGAGATTGGCGTCAAGCCACTGGCCTGCACTTCGGCCAAC
ATTTGGCCCCATGGGACCAGCCTTATGAACTACCTTCGAATGGTAGAGTATGACGGGCTGACC
GGGCGGGTTGAGTTCAACAGCAAAGGGCAGAGGACCAACTACACACTACGCATCCTGGAGAAG
TCCCGCCAGGGCCACCGTGAGATAGGGGTGTGGTACTCTAACCGGACCCTGGCCATGAATGCC
ACCACCCTGGACATCAACCTGTCACAGACTCTAGCCAACAAGACTCTGGTGGTCACAACTATC
CTGGAGAACCCGTATGTTATGCGCCGGCCCAACTTCCAGGCCTTGTCAGGGAATGAGCGCTTC
GAGGGCTTCTGCGTGGACATGCTCAGGGAGCTGGCCGAGCTGCTGCGCTTCCGATACCGCCTG
CGGTTGGTAGAGGACGGACTCTACGGGGCACCTGAGCCCAACGGTTCCTGGACAGGCATGGTT
GGAGAACTCATCAACCGGAAGGCAGACCTGGCTGTGGCAGCCTTCACCATCACCGCCGAGAGG
GAGAAGGTCATCGACTTCTCCAAGCCCTTCATGACCCTGGGGATCAGCATCCTCTACAGGGTG
CACATG**GGCCGCAAGCCTGGCTACTTCTCCTTCCTGGACCCCTTC**TCCCCTGCCGTGTGGCTC
TTCATGCTTCTTGCCTACCTGGCTGTCAGCTGTGTCTTGTTCCTGGCTGCCAG**GCTGAGCCCT**
TATGAGTGGTACAACCCACACCCGTGTCTCCGGGCGCGTCCCCATATCCTGGAGAACCAGTAC
ACGCTGGGCAACAGCCTCTGGTTCCCCGTGGGTGGCTTCATGCAGCAGGGCTCGGAGGTCATG
CCGCGGGCACTGTCCACACGCTGTGTCAGCGGAGTCTGGTGGGCCTTCACCTTGATCATCATC
TCCTCCTACACGGCCAACCTGGCTGCCTTCCTCACGGTGCAGCGCATGGAGGTGCCGGTGGAG
TCGGCTGACGACCTGGCGGATCAGACCAACATTGAGTACGGCACTATCCACGCTGGCTCCACC
ATGACCTTCTTCCAGAACTCGCGGTACCAGACGTACCAGCGGATGTGGAACTACATGCAATCG
AAGCAGCCCAGCGTGTTTGTCAAGAGCACAGAGGAGGGAATCGCCCGCGTCCTCAACTCCCGC

FIGURE 14A

TATGCCTTCCTGCTGGAGTCCACCATGAACGAGTACCACAGGCGCCTCAATTGCAACCTCACC
CAGATCGGGGGCCTCCTCGACACCAAGGGCTACGGCATCGGCATGCCGCTGGGCTCCCCTTTC
CGGGATGAGATCACACTGGCCATCCTGCAGCTCCAGGAGAACAACAGGCTGGAGATCCTGAAG
CGCAAGTGGTGGGAGGGCGGCCGGTGCCCCAAGGAGGAGGACCACAGGGCCAAAGGTTTGGGC
ATGGAGAACATTGGCGGCATTTTTGTCGTGCTGATCTGTGGCCTCATCATTGCTGTCTTCGTG
GCGGTCATGGAGTTCATCTGGTCCACGCGGAGGTCAGCGGAGTCCGAGGAGGTGTCGGTGTGC
CAGGAGATGCTGCAGGAGCTACGCCACGCCGTGTCTTGCCGAAAGACCTCGCGTTCCCGCCGG
CGCCGGCGCCCTGGTGGCCCGAGCCGGGCCCTGCTGTCGCTGCGCGCAGTCCGCGAGATGCGA
CTCAGCAACGGCAAGCTCTACTCGGCCGGCGCGGGCGGGGACGCGGGCGCGCACGGGGGTCCG
CAGCGCCTCCTGGACGACCCCGGACCTCCTGGGGGACCCCGGCCCCAGGCTCCCACGCCCTGC
ACGCACGTGCGCGTCTGCCAGGAGTGCAGGCGCATCCAGGCGCTGCGAGCTTCGGGGGCCGGG
GCGCCCCCACGTGGCCTGGGCACCCCAGCCGAAGCCACCAGCCCGCCTCGGCCGCGGCCAGGC
CCCACCGGACCCCGCGAGCTGACCGAGCACGAATGACCGTGGACGGGGCCGGGCGTGCGCCGA
CTGACTGCAGGGACGCGATGCGCCCCAGACGGACAGGACGCCGCGATTTTGCCTTCAATTCCT
GGTGAAGTCCGAGCCCGGCTCCGGAGCAGGCCTGCGCCTCCTAGTGGACTTGAGCAAGGGTGT
CGCGGACGCCGCATTCTATCCGCACCGTGGCGGAGGAGCGCAGAGACCGAGGACTCTAAGGGC
CGCGGACTGTGGGGGCTGCTCCCGGAGTTGGAAAGCGGTCCGCGGGAGGACCCCAACCTGGGA
CTGCCCAGATGCCCCAAGACTTGACACGGCTCTCCACACTTCTGGAGGTGGGGAGGGCTCTGG
ACATGTGGGTGTACCCTGTGCCCCTCTCTTCTCTTTCTCTCTTTTTTGGGGGGAGAAACCTCA
GAAATTTCTATGAGACGCCCCCAGGGAGAGGGTCTCTTGGGCCCCTATACCTCCCCCTGCCAC
ATCCCAGTCCTGTTGG

FIGURE 14B

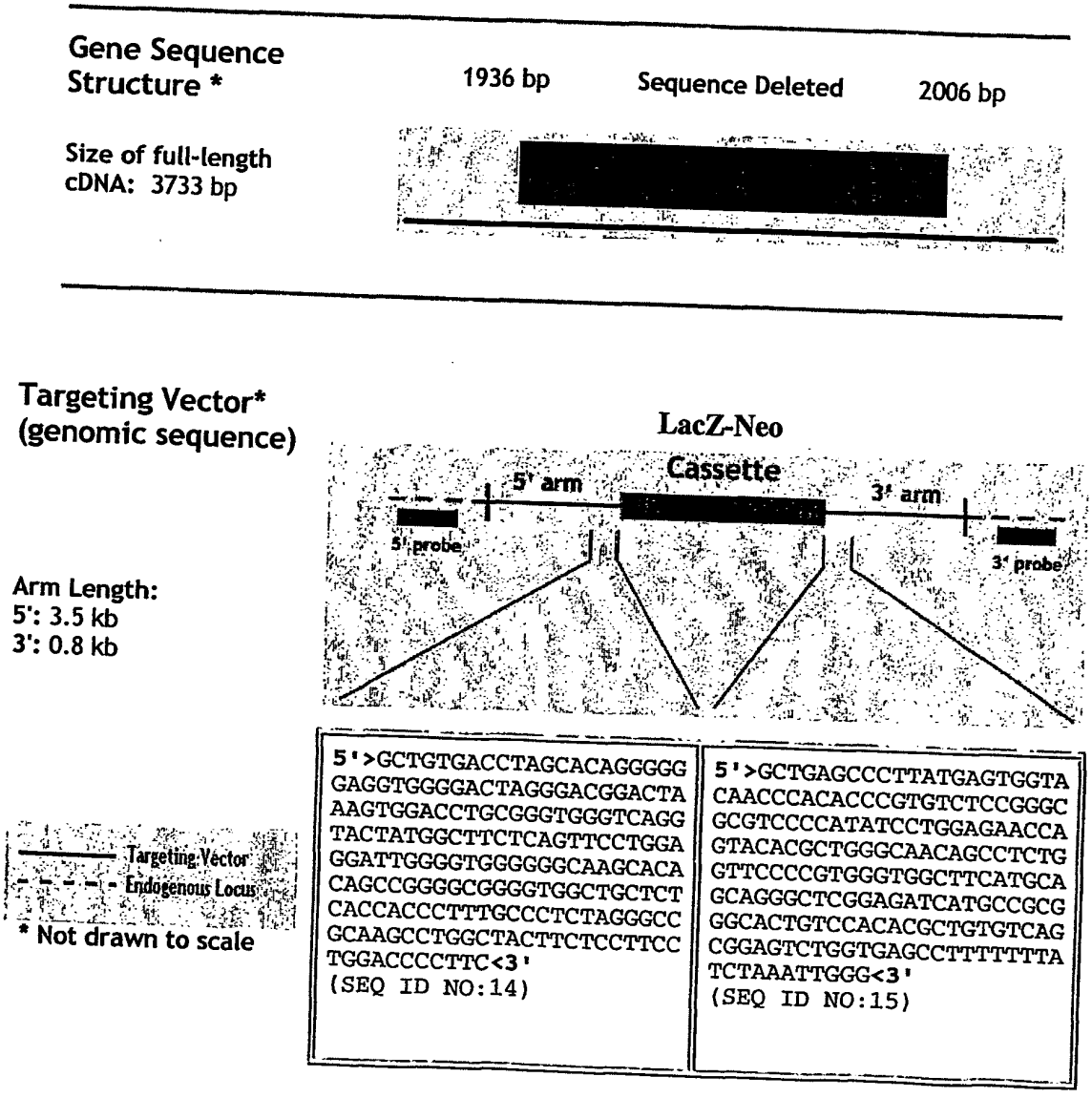

Gene Sequence Structure *
1936 bp
Sequence Deleted
2006 bp
Size of full-length cDNA: 3733 bp
Targeting Vector* (genomic sequence)
LacZ-Neo Cassette
5' arm
3' arm
5' probe
3' probe
Arm Length:
5': 3.5 kb
3': 0.8 kb
Targeting Vector
Endogenous Locus
* Not drawn to scale
5'>GCTGTGACCTAGCACAGGGGG
GAGGTGGGGACTAGGGACGGACTA
AAGTGGACCTGCGGGTGGGTCAGG
TACTATGGCTTCTCAGTTCCTGGA
GGATTGGGGTGGGGGGCAAGCACA
CAGCCGGGGCGGGGTGGCTGCTCT
CACCACCCTTTGCCCTCTAGGGCC
GCAAGCCTGGCTACTTCTCCTTCC
TGGACCCCTTC<3'
(SEQ ID NO:14)
5'>GCTGAGCCCTTATGAGTGGTA
CAACCCACACCCGTGTCTCCGGGC
GCGTCCCCATATCCTGGAGAACCA
GTACACGCTGGGCAACAGCCTCTG
GTTCCCCGTGGGTGGCTTCATGCA
GCAGGGCTCGGAGATCATGCCGCG
GGCACTGTCCACACGCTGTGTCAG
CGGAGTCTGGTGAGCCTTTTTTTA
TCTAAATTGGG<3'
(SEQ ID NO:15)

FIGURE 15

SEQUENCE LISTING

<110> Deltagen, Inc.

<120> TRANSGENIC MICE CONTAINING TARGET GENE DISRUPTIONS

<130> 13 PCT

<160> 15

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 532
<212> DNA
<213> Mus musculus

<220>
<223> Targeting vector

<400> 1
aattcggcac gcaggggttc ttctcatttc cccctacctc tctgtcacaa ccctgtcttt 60
tcctggtcag tgagacatgg gctccccatg cccggcagag aaccagctgg gaacacactt 120
cctaaaacaa gcgacaggtg agcccagtga tccctggcct gacagtctga ggcaatgcca 180
tgtcttccag agacagggac ttgcatcctg gacaccatca ctttggctcc tgcagccccт 240
tgagccagct ctggccgggc cccgagccta agtcagtcaa gggcctttac tacagcaggg 300
cccggaaggt gggcaaccag gacgcctctc cggagccaac ttgaaggaga tcctagtgaa 360
tgtgggtggc cagcggtacc tgctgccctg gagcaccctg gatgccttcc cgctgagccg 420
cctgagcagg ctccggctgt gcgccagcca tgaggagatc acgcagctct ggaatgacta 480
cgatgaggac agccaggagt tcttcttcga caggaacccc agcgccttcg gg 532

<210> 2
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 2
ggagactcac cctccctttc ttctcttctt cttcctctgt tggcagacag gtgagcccag 60
tgatccctgg cctgacagtc tgaggcaatg cccatgtctt ccagagacag ggacttgcat 120
cctggacacc atcactttgg ctcctgcagc cccttgagcc agctctggcc gggccccgag 180
cctaagtcag tcaagggcct 200

<210> 3
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 3
ccatgaggag atcacgcagc tctgcgatga ctacgatgag gacagccagg agttcttctt 60
cgacaggaac cccagcgcct tcggggtgat cgtgagcttc ctggccgcgg gaaagctggt 120
gcttctgcga gagatgtgcg ccctgtcctt ccgggaggag ctgagctact ggggcatcga 180
ggaaaccaac ctggagcgct 200

<210> 4
<211> 2469
<212> DNA
<213> Mus musculus

<400> 4
gggcgcggcg ccgtcctcgc ccgcgctcgc tccccccacc caccccggac tcccccatgt 60
atgacgactc ctacgtgccc gggtttgagg actcggaggc gggttcagcc gactcctaca 120
ccagccgccc ctctctggac tcagacgtct ccctggagga ggaccgggag agtgcccggc 180
gagaagtgga gagtcaggct cagcagcagc tggaaagagc caagcacaaa cctgtggcat 240
ttgctgtgag gaccaatgtc agctactgtg gggttctgga tgaggagtgc ccggtccagg 300
cctctggagt caactttgag gccaaagatt ttctgcacat taaagagaag tacagcaatg 360
actggtggat cgggaggcta gtgaaagaag gcggagcaat cgccttcatc cccagccccc 420
aacgcctgga gagcatccgc ctcaaacagg aacagaaggc caggagatct gggaaccctt 480
ccagcctggg tgacattggc ttccgacgct ctcctcctcc gtctctagcc aagcagaagc 540
aaaagcaggc ggaacatgtt cccccgtatg atgtggtgcc ctccatgcgg cctgtggtgc 600
tggtgggacc ctctctgaaa ggctatgagg tcacggacat gatgcagaag gcgctcttcg 660
acttccttaa acacaggttt gatggcagga tctccatcac acgcgtcacg gccgacctct 720
cgctggccaa gcgctctgtg ctcaacaatc ctggcaagag gaccatcatt gagcgctcct 780
ccgcccgctc cagcattgct gaagtgcaga gtgagattga gcgcatattc gagctggcca 840
aatccctgca gctagtggtg ttggatgctg acaccatcaa ccacccagca caacttgcca 900
agacctcact ggctcccatc atcgtctttg tcaaagtgtc ctcgccaaag gtactgcagc 960
gactgatccg atccagggga aagtcacaga tgaagcacct cactgtacag atgatggcgt 1020
acgataagct ggttcagtgc ccacccgagt cattcgacgc gattctggat gagaaccagc 1080
tagaagatgc ctgtgagcac ctggctgaat acttagaggt ttactggcga gcgacccacc 1140
acccagcacc gggccccgga ctgctgggtc cgcctagtgc catccctgga cttcagaacc 1200
agcagcagct gggggagcga gtggaggagc actcacccct ggagagggac agcctgatgc 1260
cctcagatga ggccagcgag agctcccgcc aggcctggac cggatcttca cagcgcagct 1320
ctcgccatct ggaggaggac tatgcagatg cctaccagga cctgtaccag cctcaccgcc 1380
aacacacctc ggggctgccc agtgctaacg ggcacgaccc ccaagaccgg ctcctagccc 1440
aggactcgga gcatgaccac aatgaccgga actggcagcg taaccggcct tggcccaagg 1500
acagctactg accacctcct gccccaccct ggcaggcgca ggcacagcgg ctggggtgtc 1560
cacctcaggc aggttggagt tagattggca ttaggctgcc gttagttcag ctcacacaac 1620
cctctgccca gccccaggtc cagggctgac tgtggtccca aggttctggg agaagcaagg 1680
ggcccctcac ctcctgggca cagtgacccc gtaggttctc atccgggtac tagccgtgtt 1740
ctgcatcctt ggcacctccc cctgcataag ctgccgcccc cgtgggcaac aatctcaggc 1800
caggatcact tagcaggggc cttccagcca gaatggatgc ccctctaaag agcaagaggg 1860
tgtgagtgtg ggcaacatag cctgaggaag aagaaactcg gttcctaagc aggtgtagat 1920
cctaagcaaa gggactccat tcacgccact gccacacatc agaaatgaag caatcagagc 1980
tcaacatggc ggcacttctg tcccatcagc tggggtgggc acttacacct aagacaggag 2040
cagtgcgggt gaggcaggac agacagactc acagctgtag ctctgctaga aaacggggga 2100
ctcaaccaaa ccgggaggct tagcatctgg tgagactggg gaactggggc atattcaagc 2160
caagagccag cctggactgg gggggagggt gggacagctt ccggcccccc ttgctcttct 2220
cattctttgc ccttgcatct gtcatttctg tcctttccct ccatggctcc tgcaagatag 2280
gggcttcctg actcatagca gccacttcag ttagggttag atgagaggaa caggacacag 2340
tgaacagccc ccgaggctgt ccacctggct acccttgcct tatggctcta gcgtgtgacc 2400
tacagagcat gctccattaa gaacccgccc cacctcattg tcatctccaa taaaacacca 2460
cgcacagtc 2469

<210> 5
<211> 484
<212> PRT
<213> Mus musculus

<400> 5
Met Tyr Asp Asp Ser Tyr Val Pro Gly Phe Glu Asp Ser Glu Ala Gly
1 5 10 15
Ser Ala Asp Ser Tyr Thr Ser Arg Pro Ser Leu Asp Ser Asp Val Ser
20 25 30
Leu Glu Glu Asp Arg Glu Ser Ala Arg Arg Glu Val Glu Ser Gln Ala
35 40 45

```
Gln Gln Gln Leu Glu Arg Ala Lys His Lys Pro Val Ala Phe Ala Val
    50                  55                  60
Arg Thr Asn Val Ser Tyr Cys Gly Val Leu Asp Glu Glu Cys Pro Val
65                  70                  75                  80
Gln Ala Ser Gly Val Asn Phe Glu Ala Lys Asp Phe Leu His Ile Lys
                85                  90                  95
Glu Lys Tyr Ser Asn Asp Trp Trp Ile Gly Arg Leu Val Lys Glu Gly
            100                 105                 110
Gly Ala Ile Ala Phe Ile Pro Ser Pro Gln Arg Leu Glu Ser Ile Arg
        115                 120                 125
Leu Lys Gln Glu Gln Lys Ala Arg Arg Ser Gly Asn Pro Ser Ser Leu
    130                 135                 140
Gly Asp Ile Gly Phe Arg Arg Ser Pro Pro Pro Ser Leu Ala Lys Gln
145                 150                 155                 160
Lys Gln Lys Gln Ala Glu His Val Pro Pro Tyr Asp Val Val Pro Ser
                165                 170                 175
Met Arg Pro Val Val Leu Val Gly Pro Ser Leu Lys Gly Tyr Glu Val
            180                 185                 190
Thr Asp Met Met Gln Lys Ala Leu Phe Asp Phe Leu Lys His Arg Phe
        195                 200                 205
Asp Gly Arg Ile Ser Ile Thr Arg Val Thr Ala Asp Leu Ser Leu Ala
    210                 215                 220
Lys Arg Ser Val Leu Asn Asn Pro Gly Lys Arg Thr Ile Ile Glu Arg
225                 230                 235                 240
Ser Ser Ala Arg Ser Ser Ile Ala Glu Val Gln Ser Glu Ile Glu Arg
                245                 250                 255
Ile Phe Glu Leu Ala Lys Ser Leu Gln Leu Val Val Leu Asp Ala Asp
            260                 265                 270
Thr Ile Asn His Pro Ala Gln Leu Ala Lys Thr Ser Leu Ala Pro Ile
        275                 280                 285
Ile Val Phe Val Lys Val Ser Ser Pro Lys Val Leu Gln Arg Leu Ile
    290                 295                 300
Arg Ser Arg Gly Lys Ser Gln Met Lys His Leu Thr Val Gln Met Met
305                 310                 315                 320
Ala Tyr Asp Lys Leu Val Gln Cys Pro Pro Glu Ser Phe Asp Ala Ile
                325                 330                 335
Leu Asp Glu Asn Gln Leu Glu Asp Ala Cys Glu His Leu Ala Glu Tyr
            340                 345                 350
Leu Glu Val Tyr Trp Arg Ala Thr His His Pro Ala Pro Gly Pro Gly
        355                 360                 365
Leu Leu Gly Pro Pro Ser Ala Ile Pro Gly Leu Gln Asn Gln Gln Gln
    370                 375                 380
Leu Gly Glu Arg Val Glu Glu His Ser Pro Leu Glu Arg Asp Ser Leu
385                 390                 395                 400
Met Pro Ser Asp Glu Ala Ser Glu Ser Ser Arg Gln Ala Trp Thr Gly
                405                 410                 415
Ser Ser Gln Arg Ser Ser Arg His Leu Glu Glu Asp Tyr Ala Asp Ala
            420                 425                 430
Tyr Gln Asp Leu Tyr Gln Pro His Arg Gln His Thr Ser Gly Leu Pro
        435                 440                 445
Ser Ala Asn Gly His Asp Pro Gln Asp Arg Leu Leu Ala Gln Asp Ser
    450                 455                 460
Glu His Asp His Asn Asp Arg Asn Trp Gln Arg Asn Arg Pro Trp Pro
465                 470                 475                 480
Lys Asp Ser Tyr
```

<210> 6
<211> 200
<212> DNA

<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 6
gtgtatgtct gtagcaggca gtagccttct tgtctggggc ttctgcctcc ccgtgaggta 60
ctctgaggca tctgtctctc attgccctga tgatcaaatt ccccagcaca aacctgtggc 120
atttgctgtg aggaccaatg tcagctactg tggggttctg gatgaggagt gcccggtcca 180
ggcctctgga gtcaactttg 200

<210> 7
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 7
tgactggtgg atcgggaggc tagtgaaaga gggcggagac atcgccttca tccccagccc 60
ccaacgcctg gagagcatcc ggctcaaaca gggacagaag gccaggtgag aacagtgctg 120
agattcagta cacacacaac ccaaaccccc aaacacttag cactgcgccc acagaacaca 180
gctatacatt tacacacaca 200

<210> 8
<211> 3044
<212> DNA
<213> Homo sapiens

<400> 8
cagcatcgcc ttttcctccc gtttctcctt ccactcccag ctccacatcc tcctcctatt 60
ctccccctc ccctcttcaa acccccacct tccagttccc tcacctcccc tttcggctgg 120
tcccctgggg cttgcagcaa gagggagaga gagctcctga caggattgat ggtccttccc 180
caccctgtcc tctcatccgc tccctcccca gcaggcacag acatcccccc acaaaaggca 240
ggagcccagg ctgtgtggaa acagctgctc tcagacgcct ttccatttgc tctctgctgg 300
ctaggctggg ctgtgcctct gctccctctt cctctagctg agagtgggca cctggggtac 360
cgggcccccc cacctcattc cccatgaatg ctgtgggaag tcctgagggg caggagctgc 420
acaagctggg gagtggagcc tgggacaacc ccgcctacag tggtcccccc tccccacacg 480
ggacgctgag agtctgcacc atctccagca cggggcctct ccagccccaa cccaagaagc 540
ctgaagatga accccaggag acggcataca ggacccaggt gtccagctgc tgcctccata 600
tctgtcaagg catcagagga ctttggggaa caaccctgac tgagaacaca gctgagaacc 660
gggaacttta tatcaagacc accctgaggg agctgttggt atatattgtg ttcctggtgg 720
acatctgtct actgacctat ggaatgacaa gctccagtgc ttattactac accaaagtga 780
tgtctgagct cttcttacat actccatcag acactggagt ctcctttcag gccatcagca 840
gcatggcgga cttctgggat tttgcccagg gcccactact ggacagtttg tattggacca 900
aatggtacaa caaccagagc ctgggccatg gctcccactc cttcatctac tatgagaaca 960
tgctgctggg ggttccgagg ctgcggcagc taaaggtccg caatgactcc tgtgtggtgc 1020
atgaagactt ccgggaggac attctgagct gctatgatgt ctactctcca gacaaagaag 1080
aacaactccc ctttgggccc ttcaatggca cagcgtggac ataccactcg caggatgagt 1140
tgggggggctt ctcccactgg ggcaggctca caagctacag cggaggtggc tactacctgg 1200
accttccagg atcccgacag ggtagtgcag aggctctccg ggcccttcag gaggggctgt 1260
ggctggacag gggcactcga gtggtgttca tcgacttctc agtctacaat gccaatatca 1320
atcttttctg tgtcctgagg ctggtggtgg agtttccagc tacaggaggt gccatcccat 1380
cctggcaaat ccgcacagtc aagctgatcc gctatgtcag caactgggac ttctttatcg 1440
ttggctgtga ggtcatcttc tgcgtcttca tcttctacta tgtggtggaa gagatcctgg 1500
agctccacat tcaccggctt cgctacctca gcagcatctg gaacatactg gacctggtga 1560
tcatcttgct ctccattgtg gctgtgggct tccacatatt ccgaaccctc gaggtgaatc 1620
ggctcatggg gaagctcctg cagcagccaa acacgtatgc agactttgag ttcctcgcct 1680
tctggcagac acagtacaac aacatgaatg ctgtcaacct cttcttcgcc tggatcaaga 1740
tattcaagta catcagcttc aacaaaacca tgacccagct ctcctccacg ctggcccgct 1800

gtgccaagga catcctgggc ttcgccgtca tgttcttcat tgttttcttc gcctatgccc 1860
aactcggcta cctgcttttc gggacccaag tggaaaactt tagcactttc atcaagtgca 1920
ttttcactca gttccggata atcctcgggg actttgacta caatgctatc gacaatgcca 1980
accgcatcct gggccctgcc tactttgtca cctatgtctt cttcgtcttc ttcgtgctcc 2040
tgaacatgtt cctggccatc atcaatgaca catattcaga ggtcaaggag gagctggctg 2100
gacagaagga tgagctgcaa ctttctgacc tcctgaaaca gggctacaac aagaccctac 2160
taagactgcg tctgaggaag gagagggttt cggatgtgca gaaggtcctg cagggtgggg 2220
agcaggagat ccagtttgag gatttcacca acaccttaag ggaactggga cacgcagagc 2280
atgaaatcac tgagctcacg gccaccttca ccaagtttga cagagatggg aatcgtattc 2340
tggatgagaa ggaacaggaa aaaatgcgac aggacctgga ggaagagagg gtggccctca 2400
acactgagat tgagaaacta ggccgatcta ttgtgagcag cccacaaggc aaatcgggtc 2460
cagaggctgc cagagcagga ggctgggttt caggagaaga attctacatg ctcacaagga 2520
gagttctgca gctggagact gtcctggaag gagtagtgtc ccagattgat gctgtaggct 2580
caaagctgaa aatgctggag aggaaggggt ggctggctcc ctccccaggc gtgaaggaac 2640
aagctatttg gaagcacccg cagccagccc cagctgtgac cccagacccc tggggagtcc 2700
agggtgggca ggagagtgag gttccctata aaagagaaga ggaagcctta gaggagagga 2760
gactctcccg tggtgagatt ccaacgttgc agaggagtta agtgtgaggc actcccggag 2820
caaagtctat gaaggatctt ctgcaagagg ctgcctcctg gtccactgaa cctggaaact 2880
gagtgggctt taaccaggag ataaaaatgg agcctgaagg gaatcaggca aggaaatgaa 2940
ctcaggattc agagatcttt gaattaatat gtggtgggtt ctgacattat tcttccataa 3000
gaccatgtgg gtttccatgg tggctatcaa taaaactcct tagt 3044

<210> 9
<211> 805
<212> PRT
<213> Homo sapiens

<400> 9

```
Met Asn Ala Val Gly Ser Pro Glu Gly Gln Glu Leu His Lys Leu Gly
 1               5                  10                  15
Ser Gly Ala Trp Asp Asn Pro Ala Tyr Ser Gly Pro Pro Ser Pro His
            20                  25                  30
Gly Thr Leu Arg Val Cys Thr Ile Ser Ser Thr Gly Pro Leu Gln Pro
        35                  40                  45
Gln Pro Lys Lys Pro Glu Asp Glu Pro Gln Glu Thr Ala Tyr Arg Thr
    50                  55                  60
Gln Val Ser Ser Cys Cys Leu His Ile Cys Gln Gly Ile Arg Gly Leu
65                  70                  75                  80
Trp Gly Thr Thr Leu Thr Glu Asn Thr Ala Glu Asn Arg Glu Leu Tyr
                85                  90                  95
Ile Lys Thr Thr Leu Arg Glu Leu Leu Val Tyr Ile Val Phe Leu Val
            100                 105                 110
Asp Ile Cys Leu Leu Thr Tyr Gly Met Thr Ser Ser Ser Ala Tyr Tyr
        115                 120                 125
Tyr Thr Lys Val Met Ser Glu Leu Phe Leu His Thr Pro Ser Asp Thr
    130                 135                 140
Gly Val Ser Phe Gln Ala Ile Ser Ser Met Ala Asp Phe Trp Asp Phe
145                 150                 155                 160
Ala Gln Gly Pro Leu Leu Asp Ser Leu Tyr Trp Thr Lys Trp Tyr Asn
                165                 170                 175
Asn Gln Ser Leu Gly His Gly Ser His Ser Phe Ile Tyr Tyr Glu Asn
            180                 185                 190
Met Leu Leu Gly Val Pro Arg Leu Arg Gln Leu Lys Val Arg Asn Asp
        195                 200                 205
Ser Cys Val Val His Glu Asp Phe Arg Glu Asp Ile Leu Ser Cys Tyr
    210                 215                 220
Asp Val Tyr Ser Pro Asp Lys Glu Glu Gln Leu Pro Phe Gly Pro Phe
225                 230                 235                 240
Asn Gly Thr Ala Trp Thr Tyr His Ser Gln Asp Glu Leu Gly Gly Phe
                245                 250                 255
Ser His Trp Gly Arg Leu Thr Ser Tyr Ser Gly Gly Gly Tyr Tyr Leu
```

```
            260                 265                 270
Asp Leu Pro Gly Ser Arg Gln Gly Ser Ala Glu Ala Leu Arg Ala Leu
        275                 280                 285
Gln Glu Gly Leu Trp Leu Asp Arg Gly Thr Arg Val Val Phe Ile Asp
    290                 295                 300
Phe Ser Val Tyr Asn Ala Asn Ile Asn Leu Phe Cys Val Leu Arg Leu
305                 310                 315                 320
Val Val Glu Phe Pro Ala Thr Gly Gly Ala Ile Pro Ser Trp Gln Ile
                325                 330                 335
Arg Thr Val Lys Leu Ile Arg Tyr Val Ser Asn Trp Asp Phe Phe Ile
            340                 345                 350
Val Gly Cys Glu Val Ile Phe Cys Val Phe Ile Phe Tyr Tyr Val Val
        355                 360                 365
Glu Glu Ile Leu Glu Leu His Ile His Arg Leu Arg Tyr Leu Ser Ser
    370                 375                 380
Ile Trp Asn Ile Leu Asp Leu Val Ile Ile Leu Leu Ser Ile Val Ala
385                 390                 395                 400
Val Gly Phe His Ile Phe Arg Thr Leu Glu Val Asn Arg Leu Met Gly
                405                 410                 415
Lys Leu Leu Gln Gln Pro Asn Thr Tyr Ala Asp Phe Glu Phe Leu Ala
            420                 425                 430
Phe Trp Gln Thr Gln Tyr Asn Asn Met Asn Ala Val Asn Leu Phe Phe
        435                 440                 445
Ala Trp Ile Lys Ile Phe Lys Tyr Ile Ser Phe Asn Lys Thr Met Thr
    450                 455                 460
Gln Leu Ser Ser Thr Leu Ala Arg Cys Ala Lys Asp Ile Leu Gly Phe
465                 470                 475                 480
Ala Val Met Phe Phe Ile Val Phe Phe Ala Tyr Ala Gln Leu Gly Tyr
                485                 490                 495
Leu Leu Phe Gly Thr Gln Val Glu Asn Phe Ser Thr Phe Ile Lys Cys
            500                 505                 510
Ile Phe Thr Gln Phe Arg Ile Ile Leu Gly Asp Phe Asp Tyr Asn Ala
        515                 520                 525
Ile Asp Asn Ala Asn Arg Ile Leu Gly Pro Ala Tyr Phe Val Thr Tyr
    530                 535                 540
Val Phe Phe Val Phe Phe Val Leu Leu Asn Met Phe Leu Ala Ile Ile
545                 550                 555                 560
Asn Asp Thr Tyr Ser Glu Val Lys Glu Glu Leu Ala Gly Gln Lys Asp
                565                 570                 575
Glu Leu Gln Leu Ser Asp Leu Leu Lys Gln Gly Tyr Asn Lys Thr Leu
            580                 585                 590
Leu Arg Leu Arg Leu Arg Lys Glu Arg Val Ser Asp Val Gln Lys Val
        595                 600                 605
Leu Gln Gly Gly Glu Gln Glu Ile Gln Phe Glu Asp Phe Thr Asn Thr
    610                 615                 620
Leu Arg Glu Leu Gly His Ala Glu His Glu Ile Thr Glu Leu Thr Ala
625                 630                 635                 640
Thr Phe Thr Lys Phe Asp Arg Asp Gly Asn Arg Ile Leu Asp Glu Lys
                645                 650                 655
Glu Gln Glu Lys Met Arg Gln Asp Leu Glu Glu Glu Arg Val Ala Leu
            660                 665                 670
Asn Thr Glu Ile Glu Lys Leu Gly Arg Ser Ile Val Ser Ser Pro Gln
        675                 680                 685
Gly Lys Ser Gly Pro Glu Ala Ala Arg Ala Gly Gly Trp Val Ser Gly
    690                 695                 700
Glu Glu Phe Tyr Met Leu Thr Arg Arg Val Leu Gln Leu Glu Thr Val
705                 710                 715                 720
Leu Glu Gly Val Val Ser Gln Ile Asp Ala Val Gly Ser Lys Leu Lys
                725                 730                 735
Met Leu Glu Arg Lys Gly Trp Leu Ala Pro Ser Pro Gly Val Lys Glu
            740                 745                 750
```

```
Gln Ala Ile Trp Lys His Pro Gln Pro Ala Pro Ala Val Thr Pro Asp
        755                 760                 765
Pro Trp Gly Val Gln Gly Gly Gln Glu Ser Glu Val Pro Tyr Lys Arg
    770                 775                 780
Glu Glu Glu Ala Leu Glu Glu Arg Arg Leu Ser Arg Gly Glu Ile Pro
785                 790                 795                 800
Thr Leu Gln Arg Ser
                805
```

<210> 10
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 10
tccttatttg gttaactgaa ggagacggag acgcaggcat ggaattgggg tcgatggtaa 60
agaagagagc acatctgtac cttcacattg aaaagctgat gacttcagtg attgctgggc 120
ccttttatag gggaggtcca tgccacatgc tgggatagct tgatgtttag aacgcgtcgc 180
tgtagtatct gagaatgagg 200

<210> 11
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 11
gttcctggcc atcatcaacg acacatactc cgaggtcaag gaggagctgg ctggccagaa 60
ggatcagttg cagctttctg acttcctgaa acaggtgact acgcaggctc tgcatagcat 120
ctctacactg gggtttggtt ctaaggtgct gtgccacggt ggactgttct tgattggtgc 180
aggcgcctgt gctcttgaat 200

<210> 12
<211> 3733
<212> DNA
<213> Mus musculus

<400> 12
ccccttctct cctccatttt ctcctttccc cttcctcccc tcctggcccg ccctttgctt 60
ttccacctgt ccttcctccc cgctgctggg gagacaccgt tttggattag ttgggggcca 120
ccgccggcgg cgggggaggg ggcccagtgg actgccctct cccccggccc cagcctcacc 180
agcacccagc atcggagcac ctcctggttg tcgttcctcg ggcctcgagg gagcccagcc 240
ctccgtccca ccaggatccg tggcgagtgg gggccgcggc agctgcgtcc ccatgaggag 300
gggaggaaga tgccggctga gctgctgctg ctgctgatag tcgccttcgc caatcccagc 360
tgccaggtgc tgtcatcact gcgcatggct gcaatcctgg acgaccagac cgtgtgtggc 420
cgtggtgagc gtctggccct ggccctggcc cgagagcaga tcaatgggat catcgaggtc 480
ccagccaagg ccagagtgga agtagacatc tttgagctgc agcgggacag ccagtacgag 540
accacggaca ccatgtgtca gatcctgccc aaggggttg tatctgtctt gggaccctcc 600
tccagcccag cttctgcctc caccgtgagc catatctgtg gggagaagga gattccccac 660
atcaaggtgg gtcctgagga gacgccccgc cttcagtacc ttcgcttcgc atctgtcagc 720
ctgtacccca gtaatgaaga tgtcagcctg gcagtctccc gaatcctcaa gtcctttaac 780
taccctcag ctagcctcat ctgcgccaag gctgagtgcc tgctgcggct agaagaactg 840
gtgcgaggct tcctcatctc caaggagaca ctgtccgtga ggatgcttga tgacagccgg 900
gaccccacgc cgctactcaa ggagatccga gatgacaaag tgtccaccat catcattgat 960
gccaatgcgt ccatctccca ccttgtcctc cgtaaggctt cggagctggg aatgacctca 1020

gcgttttaca agtacatcct caccaccatg gactttccca tcctgcatct ggatggtatc 1080
gtggaggact cctccaacat cctgggcttt tccatgttca acacctccca ccccttctac 1140
ccagagtttg tgcgcagcct caacatgtcc tggagggaga actgtgaagc cagcacctat 1200
cctggccctg cgctgtccgc agccctgatg tttgacgctg tgcacgtggt ggtaagcgct 1260
gtccgagaac tgaaccgaag ccaggagatt ggcgtcaagc cactggcctg cacttcggcc 1320
aacatttggc cccatgggac cagccttatg aactaccttc gaatggtaga gtatgacggg 1380
ctgaccgggc gggttgagtt caacagcaaa gggcagagga ccaactacac actacgcatc 1440
ctggagaagt cccgccaggg ccaccgtgag ataggggtgt ggtactctaa ccggaccctg 1500
gccatgaatg ccaccaccct ggacatcaac ctgtcacaga ctctagccaa caagactctg 1560
gtggtcacaa ctatcctgga gaacccgtat gttatgcgcc ggcccaactt ccaggccttg 1620
tcagggaatg agcgcttcga gggcttctgc gtggacatgc tcagggagct ggccgagctg 1680
ctgcgcttcc gataccgcct gcggttggta gaggacggac tctacggggc acctgagccc 1740
aacggttcct ggacaggcat ggttggagaa ctcatcaacc ggaaggcaga cctggctgtg 1800
gcagccttca ccatcaccgc cgagagggag aaggtcatcg acttctccaa gcccttcatg 1860
accctgggga tcagcatcct ctacagggtg cacatgggcc gcaagcctgg ctacttctcc 1920
ttcctggacc ccttctcccc tgccgtgtgg ctcttcatgc ttcttgccta cctggctgtc 1980
agctgtgtct tgttcctggc tgccaggctg agcccttatg agtggtacaa cccacacccg 2040
tgtctccggg cgcgtcccca tatcctggag aaccagtaca cgctgggcaa cagcctctgg 2100
ttccccgtgg gtggcttcat gcagcagggc tcggaggtca tgccgcgggc actgtccaca 2160
cgctgtgtca gcggagtctg gtgggccttc accttgatca tcatctcctc ctacacggcc 2220
aacctggctg ccttcctcac ggtgcagcgc atggaggtgc cggtggagtc ggctgacgac 2280
ctggcggatc agaccaacat tgagtacggc actatccacg ctggctccac catgaccttc 2340
ttccagaact cgcggtacca gacgtaccag cggatgtgga actacatgca atcgaagcag 2400
cccagcgtgt ttgtcaagag cacagaggag ggaatcgccc gcgtcctcaa ctcccgctat 2460
gccttcctgc tggagtccac catgaacgag taccacaggc gcctcaattg caacctcacc 2520
cagatcgggg gcctcctcga caccaagggc tacggcatcg gcatgccgct gggctcccct 2580
ttccgggatg agatcacact ggccatcctg cagctccagg agaacaacag gctggagatc 2640
ctgaagcgca agtggtggga gggcggccgg tgccccaagg aggaggacca cagggccaaa 2700
ggtttgggca tggagaacat tggcggcatt tttgtcgtgc tgatctgtgg cctcatcatt 2760
gctgtcttcg tggcggtcat ggagttcatc tggtccacgc ggaggtcagc ggagtccgag 2820
gaggtgtcgg tgtgccagga gatgctgcag gagctacgcc acgccgtgtc ttgccgaaag 2880
acctcgcgtt cccgccggcg ccggcgccct ggtggcccga gccgggccct gctgtcgctg 2940
cgcgcagtcc gcgagatgcg actcagcaac ggcaagctct actcggccgg cgcgggcggg 3000
gacgcgggcg cgcacggggg tccgcagcgc ctcctggacg accccggacc tcctgggga 3060
ccccggcccc aggctcccac gccctgcacg cacgtgcgcg tctgccagga gtgcaggcgc 3120
atccaggcgc tgcgagcttc gggggccggg gcgcccccac gtggcctggg caccccagcc 3180
gaagccacca gcccgcctcg gccgcggcca ggccccaccg gaccccgcga gctgaccgag 3240
cacgaatgac cgtggacggg gccgggcgtg cgccgactga ctgcagggac gcgatgcgcc 3300
ccagacggac aggacgccgc gattttgcct tcaattcctg gtgaagtccg agcccggctc 3360
cggagcaggc ctgcgcctcc tagtggactt gagcaagggt gtcgcggacg ccgcattcta 3420
tccgcaccgt ggcggaggag cgcagagacc gaggactcta agggccgcgg actgtggggg 3480
ctgctcccgg agttggaaag cggtccgcgg gaggacccca acctgggact gcccagatgc 3540
cccaagactt gacacggctc tccacacttc tggaggtggg gagggctctg gacatgtggg 3600
tgtaccctgt gcccctctct tctctttctc tctttttttgg ggggagaaac ctcagaaatt 3660
tctatgagac gcccccaggg agagggtctc ttgggcccct atacctcccc ctgccacatc 3720
ccagtcctgt tgg 3733

<210> 13
<211> 979
<212> PRT
<213> Mus musculus

<400> 13
Met Pro Ala Glu Leu Leu Leu Leu Leu Ile Val Ala Phe Ala Asn Pro
1 5 10 15
Ser Cys Gln Val Leu Ser Ser Leu Arg Met Ala Ala Ile Leu Asp Asp
20 25 30
Gln Thr Val Cys Gly Arg Gly Glu Arg Leu Ala Leu Ala Leu Ala Arg
35 40 45
Glu Gln Ile Asn Gly Ile Ile Glu Val Pro Ala Lys Ala Arg Val Glu
50 55 60

```
Val Asp Ile Phe Glu Leu Gln Arg Asp Ser Gln Tyr Glu Thr Thr Asp
65                  70                  75                  80
Thr Met Cys Gln Ile Leu Pro Lys Gly Val Val Ser Val Leu Gly Pro
                85                  90                  95
Ser Ser Ser Pro Ala Ser Ala Ser Thr Val Ser His Ile Cys Gly Glu
            100                 105                 110
Lys Glu Ile Pro His Ile Lys Val Gly Pro Glu Glu Thr Pro Arg Leu
        115                 120                 125
Gln Tyr Leu Arg Phe Ala Ser Val Ser Leu Tyr Pro Ser Asn Glu Asp
    130                 135                 140
Val Ser Leu Ala Val Ser Arg Ile Leu Lys Ser Phe Asn Tyr Pro Ser
145                 150                 155                 160
Ala Ser Leu Ile Cys Ala Lys Ala Glu Cys Leu Leu Arg Leu Glu Glu
                165                 170                 175
Leu Val Arg Gly Phe Leu Ile Ser Lys Glu Thr Leu Ser Val Arg Met
            180                 185                 190
Leu Asp Asp Ser Arg Asp Pro Thr Pro Leu Leu Lys Glu Ile Arg Asp
        195                 200                 205
Asp Lys Val Ser Thr Ile Ile Ile Asp Ala Asn Ala Ser Ile Ser His
    210                 215                 220
Leu Val Leu Arg Lys Ala Ser Glu Leu Gly Met Thr Ser Ala Phe Tyr
225                 230                 235                 240
Lys Tyr Ile Leu Thr Thr Met Asp Phe Pro Ile Leu His Leu Asp Gly
                245                 250                 255
Ile Val Glu Asp Ser Ser Asn Ile Leu Gly Phe Ser Met Phe Asn Thr
            260                 265                 270
Ser His Pro Phe Tyr Pro Glu Phe Val Arg Ser Leu Asn Met Ser Trp
        275                 280                 285
Arg Glu Asn Cys Glu Ala Ser Thr Tyr Pro Gly Pro Ala Leu Ser Ala
    290                 295                 300
Ala Leu Met Phe Asp Ala Val His Val Val Val Ser Ala Val Arg Glu
305                 310                 315                 320
Leu Asn Arg Ser Gln Glu Ile Gly Val Lys Pro Leu Ala Cys Thr Ser
                325                 330                 335
Ala Asn Ile Trp Pro His Gly Thr Ser Leu Met Asn Tyr Leu Arg Met
            340                 345                 350
Val Glu Tyr Asp Gly Leu Thr Gly Arg Val Glu Phe Asn Ser Lys Gly
        355                 360                 365
Gln Arg Thr Asn Tyr Thr Leu Arg Ile Leu Glu Lys Ser Arg Gln Gly
    370                 375                 380
His Arg Glu Ile Gly Val Trp Tyr Ser Asn Arg Thr Leu Ala Met Asn
385                 390                 395                 400
Ala Thr Thr Leu Asp Ile Asn Leu Ser Gln Thr Leu Ala Asn Lys Thr
                405                 410                 415
Leu Val Val Thr Thr Ile Leu Glu Asn Pro Tyr Val Met Arg Arg Pro
            420                 425                 430
Asn Phe Gln Ala Leu Ser Gly Asn Glu Arg Phe Glu Gly Phe Cys Val
        435                 440                 445
Asp Met Leu Arg Glu Leu Ala Glu Leu Leu Arg Phe Arg Tyr Arg Leu
    450                 455                 460
Arg Leu Val Glu Asp Gly Leu Tyr Gly Ala Pro Glu Pro Asn Gly Ser
465                 470                 475                 480
Trp Thr Gly Met Val Gly Glu Leu Ile Asn Arg Lys Ala Asp Leu Ala
                485                 490                 495
Val Ala Ala Phe Thr Ile Thr Ala Glu Arg Glu Lys Val Ile Asp Phe
            500                 505                 510
Ser Lys Pro Phe Met Thr Leu Gly Ile Ser Ile Leu Tyr Arg Val His
        515                 520                 525
Met Gly Arg Lys Pro Gly Tyr Phe Ser Phe Leu Asp Pro Phe Ser Pro
    530                 535                 540
Ala Val Trp Leu Phe Met Leu Leu Ala Tyr Leu Ala Val Ser Cys Val
```

545 550 555 560
Leu Phe Leu Ala Ala Arg Leu Ser Pro Tyr Glu Trp Tyr Asn Pro His
565 570 575
Pro Cys Leu Arg Ala Arg Pro His Ile Leu Glu Asn Gln Tyr Thr Leu
580 585 590
Gly Asn Ser Leu Trp Phe Pro Val Gly Gly Phe Met Gln Gln Gly Ser
595 600 605
Glu Val Met Pro Arg Ala Leu Ser Thr Arg Cys Val Ser Gly Val Trp
610 615 620
Trp Ala Phe Thr Leu Ile Ile Ile Ser Ser Tyr Thr Ala Asn Leu Ala
625 630 635 640
Ala Phe Leu Thr Val Gln Arg Met Glu Val Pro Val Glu Ser Ala Asp
645 650 655
Asp Leu Ala Asp Gln Thr Asn Ile Glu Tyr Gly Thr Ile His Ala Gly
660 665 670
Ser Thr Met Thr Phe Phe Gln Asn Ser Arg Tyr Gln Thr Tyr Gln Arg
675 680 685
Met Trp Asn Tyr Met Gln Ser Lys Gln Pro Ser Val Phe Val Lys Ser
690 695 700
Thr Glu Glu Gly Ile Ala Arg Val Leu Asn Ser Arg Tyr Ala Phe Leu
705 710 715 720
Leu Glu Ser Thr Met Asn Glu Tyr His Arg Arg Leu Asn Cys Asn Leu
725 730 735
Thr Gln Ile Gly Gly Leu Leu Asp Thr Lys Gly Tyr Gly Ile Gly Met
740 745 750
Pro Leu Gly Ser Pro Phe Arg Asp Glu Ile Thr Leu Ala Ile Leu Gln
755 760 765
Leu Gln Glu Asn Asn Arg Leu Glu Ile Leu Lys Arg Lys Trp Trp Glu
770 775 780
Gly Gly Arg Cys Pro Lys Glu Glu Asp His Arg Ala Lys Gly Leu Gly
785 790 795 800
Met Glu Asn Ile Gly Gly Ile Phe Val Val Leu Ile Cys Gly Leu Ile
805 810 815
Ile Ala Val Phe Val Ala Val Met Glu Phe Ile Trp Ser Thr Arg Arg
820 825 830
Ser Ala Glu Ser Glu Glu Val Ser Val Cys Gln Glu Met Leu Gln Glu
835 840 845
Leu Arg His Ala Val Ser Cys Arg Lys Thr Ser Arg Ser Arg Arg Arg
850 855 860
Arg Arg Pro Gly Gly Pro Ser Arg Ala Leu Leu Ser Leu Arg Ala Val
865 870 875 880
Arg Glu Met Arg Leu Ser Asn Gly Lys Leu Tyr Ser Ala Gly Ala Gly
885 890 895
Gly Asp Ala Gly Ala His Gly Gly Pro Gln Arg Leu Leu Asp Asp Pro
900 905 910
Gly Pro Pro Gly Gly Pro Arg Pro Gln Ala Pro Thr Pro Cys Thr His
915 920 925
Val Arg Val Cys Gln Glu Cys Arg Arg Ile Gln Ala Leu Arg Ala Ser
930 935 940
Gly Ala Gly Ala Pro Pro Arg Gly Leu Gly Thr Pro Ala Glu Ala Thr
945 950 955 960
Ser Pro Pro Arg Pro Arg Pro Gly Pro Thr Gly Pro Arg Glu Leu Thr
965 970 975
Glu His Glu

<210> 14
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 14
gctgtgacct agcacagggg ggaggtgggg actagggacg gactaaagtg gacctgcggg 60
tgggtcaggt actatggctt ctcagttcct ggaggattgg ggtgggggggc aagcacacag 120
ccggggcggg gtggctgctc tcaccaccct ttgccctcta gggccgcaag cctggctact 180
tctccttcct ggacccccttc 200

<210> 15
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 15
gctgagccct tatgagtggt acaacccaca cccgtgtctc cgggcgcgtc cccatatcct 60
ggagaaccag tacacgctgg gcaacagcct ctggttcccc gtgggtggct tcatgcagca 120
gggctcggag atcatgccgc gggcactgtc cacacgctgt gtcagcggag tctggtgagc 180
ctttttttat ctaaattggg 200